Horizontal Gene Transfer as an Indispensable Driver for Evolution of Neocallimastigomycota into a Distinct Gut-Dwelling Fungal Lineage

Murphy, Chelsea L.; Youssef, Noha H.; Hanafy, Radwa A.; Couger, M. B.; Stajich, Jason E.; Wang, Yan; Baker, Kristina; Dagar, Sumit S.; Griffith, Gareth W.; Farag, Ibrahim F.; Callaghan, T. M.; Elshahed, Mostafa S

Published in:
Applied and Environmental Microbiology

DOI:
10.1128/AEM.00988-19

Publication date:
2019

Citation for published version (APA):
Murphy, C. L., Youssef, N. H., Hanafy, R. A., Couger, M. B., Stajich, J. E., Wang, Y., Baker, K., Dagar, S. S., Griffith, G. W., Farag, I. F., Callaghan, T. M., & Elshahed, M. S. (2019). Horizontal Gene Transfer as an Indispensable Driver for Evolution of Neocallimastigomycota into a Distinct Gut-Dwelling Fungal Lineage. *Applied and Environmental Microbiology*, *85*(15), [e00988-19]. https://doi.org/10.1128/AEM.00988-19

Horizontal gene transfer as an indispensable driver for Neocallimastigomycota evolution into a distinct gut-dwelling fungal lineage

Chelsea L. Murphy[1¶], Noha H. Youssef[1¶], Radwa A. Hanafy[1], MB Couger[2], Jason E. Stajich[3], Y. Wang[3], Kristina Baker[1], Sumit S. Dagar[4], Gareth W. Griffith[5], Ibrahim F. Farag[1], TM Callaghan[6], and Mostafa S. Elshahed[1*]

[1]Department of Microbiology and Molecular Genetics and [2] High Performance Computing Center, Oklahoma State University, Stillwater, OK. [3]Department of Microbiology and Plant Pathology, Institute for Integrative Genome Biology, University of California-Riverside, Riverside, CA. Bioenergy group, Agharkar Research Institute, Pune, India. [5]Institute of Biological, Environmental, and Rural Sciences (IBERS) Aberystwyth University, Aberystwyth, Wales, UK. [6]Department for Quality Assurance and Analytics, Bavarian State Research Center for Agriculture, Freising, Germany.

Running Title: Horizontal gene transfer in the Neocallimastigomycota

[*]Corresponding author: Mailing address: Oklahoma State University, Department of Microbiology and Molecular Genetics, 1110 S Innovation Way, Stillwater, OK 74074. Phone: (405) 744-3005, Fax: (405) 744-1112. Email: **Mostafa@okstate.edu**.

[¶]Both authors contributed equally to this work.

Abstract

Survival and growth of the anaerobic gut fungi (AGF, Neocallimastigomycota) in the herbivorous gut necessitate the possession of multiple abilities absent in other fungal lineages. We hypothesized that horizontal gene transfer (HGT) was instrumental in forging the evolution of AGF into a phylogenetically distinct gut-dwelling fungal lineage. Patterns of HGT were evaluated in the transcriptomes of 27 AGF strains, 22 of which were isolated and sequenced in this study, and 4 AGF genomes broadly covering the breadth of AGF diversity. We identified 277 distinct incidents of HGT in AGF transcriptomes, with subsequent gene duplication resulting in an HGT frequency of 2-3.5% in AGF genomes. The majority of HGT events were AGF specific (91.7%) and wide (70.8%), indicating their occurrence at early stages of AGF evolution. The acquired genes allowed AGF to expand their substrate utilization range, provided new venues for electron disposal, augmented their biosynthetic capabilities, and facilitated their adaptation to anaerobiosis. The majority of donors were anaerobic fermentative bacteria prevalent in the herbivorous gut. This work strongly indicates that HGT indispensably forged the evolution of AGF as a distinct fungal phylum and provides a unique example of the role of HGT in shaping the evolution of a high rank taxonomic eukaryotic lineage.

40 **Importance**

41 The anaerobic gut fungi (AGF) represent a distinct basal phylum lineage
42 (Neocallimastigomycota) commonly encountered in the rumen and alimentary tracts of
43 herbivores. Survival and growth of anaerobic gut fungi in these anaerobic, eutrophic, and
44 prokaryotes dominated habitats necessitates the acquisition of several traits absent in other fungal
45 lineages. This manuscript assesses the role of horizontal gene transfer as a relatively fast
46 mechanism for trait acquisition by the Neocallimastigomycota post-sequestration in the
47 herbivorous gut. Analysis of twenty-seven transcriptomes that represent the broad
48 Neocallimastigomycota diversity identified 277 distinct HGT events, with subsequent gene
49 duplication resulting in an HGT frequency of 2-3.5% in AGF genomes. These HGT events have
50 allowed AGF to survive in the herbivorous gut by expanding their substrate utilization range,
51 augmenting their biosynthetic pathway, providing new routes for electron disposal by expanding
52 fermentative capacities, and facilitating their adaptation to anaerobiosis. HGT in the AGF is also
53 shown to be mainly a cross-kingdom affair, with the majority of donors belonging to the bacteria.
54 This work represents a unique example of the role of HGT in shaping the evolution of a high
55 rank taxonomic eukaryotic lineage.
56

Introduction

Horizontal gene transfer (HGT) is defined as the acquisition, integration, and retention of foreign genetic material into a recipient organism (1). HGT represents a relatively rapid process for trait acquisition; as opposed to gene creation either from preexisting genes (via duplication, fission, fusion, or exon shuffling) or through *de-novo* gene birth from non-coding sequences (2-6). In prokaryotes, the occurrence, patterns, frequency, and impact of HGT on the genomic architecture (7), metabolic abilities (8, 9), physiological preferences (10, 11), and ecological fitness (12) has been widely investigated, and the process is now regarded as a major driver of genome evolution in bacteria and archaea (13, 14). Although eukaryotes are perceived to evolve principally through modifying existing genetic information, analysis of HGT events in eukaryotic genomes has been eliciting increasing interest and scrutiny. In spite of additional barriers that need to be overcome in eukaryotes, e.g. crossing the nuclear membrane, germline sequestration in sexual multicellular eukaryotes, and epigenetic nucleic acids modifications mechanisms (5, 15), it is now widely accepted that HGT contributes significantly to eukaryotic genome evolution (16, 17). HGT events have convincingly been documented in multiple phylogenetically disparate eukaryotes ranging from the Excavata (18-21), SAR supergroup (22-25), Algae (26), Plants (27), and Opisthokonta (28-31). Reported HGT frequency in eukaryotic genomes ranges from a handful of genes, e.g. (32), to up to 9.6% in Bdelloid rotifers (30).

The kingdom Fungi represents a phylogenetically coherent clade that evolved ≈ 900-1481 Mya from a unicellular flagellated ancestor (33-35). To date, multiple efforts have been reported on the detection and quantification of HGT in fungi. A survey of 60 fungal genomes reported HGT frequencies of 0-0.38% (29), and similar low values were observed in the genomes of five early-diverging pathogenic Microsporidia and Cryptomycota (36). A recent study has

documented the role of HGT in expanding the catabolic capabilities of members of the mycotrophic genus *Trichoderma* by extensive acquisition of plant biomass degradation capacities from plant-associated filamentous Ascomycetes (37). The osmotrophic lifestyle of fungi (38) has typically been regarded as less conducive to HGT compared to the phagocytic lifestyle of several microeukaryotes with relatively higher HGT frequency (39).

The anaerobic gut fungi (AGF, Neocallimastigomycota) represent a phylogenetically distinct basal fungal lineage. The AGF appear to exhibit a restricted distribution pattern, being encountered in the gut of ruminant and non-ruminant herbivorous (40). In the herbivorous gut, the life cycle of the AGF (Figure S1) involves the discharge of motile flagellated zoospores from sporangia in response to animal feeding, the chemotaxis and attachment of zoospores to ingested plant material, spore encystment, and the subsequent production of rhizoidal growth that penetrates and digests plant biomass through the production of a wide array of cellulolytic and lignocellulolytic enzymes.

Survival, colonization, and successful propagation of AGF in the herbivorous gut necessitate the acquisition of multiple unique physiological characteristics and metabolic abilities absent in other fungal lineages. These include, but are not limited to, development of a robust plant biomass degradation machinery, adaptation to anaerobiosis, and exclusive dependence on fermentation for energy generation and recycling of electron carriers (41, 42). Therefore, we hypothesized that sequestration into the herbivorous gut was conducive to the broad adoption of HGT as a relatively faster adaptive evolutionary strategy for niche adaptation by the AGF (Figure S1). Further, since no part of the AGF life cycle occurs outside the animal host and no reservoir of AGF outside the herbivorous gut has been identified (40), then acquisition would mainly occur from donors that are prevalent in the herbivorous gut (Figure S1). Apart from

earlier observations on the putative bacterial origin of a few catabolic genes in two AGF isolates (43, 44), and preliminary BLAST-based queries of a few genomes (42, 45), little is currently known on the patterns, determinants, and frequency of HGT in the Neocallimastigomycota. To address this hypothesis, we systematically evaluated the patterns of HGT acquisition in the transcriptomes of 27 AGF strains and 4 AGF genomes broadly covering the breadth of AGF genus-level diversity. Our results document the high level of HGT in AGF in contrast to HGT paucity across the fungal kingdom. The identity of genes transferred, distribution pattern of events across AGF genera, phylogenetic affiliation of donors, and the expansion of acquired genetic material in AGF genomes highlight the role played by HGT in forging the evolution and diversification of the Neocallimastigomycota as a phylogenetically, metabolically, and ecologically distinct lineage in the fungal kingdom.

Materials and Methods

Organisms. Type strains of the Neocallimastigomycota are unavailable through culture collections due to their strict anaerobic and fastidious nature, as well as the frequent occurrence of senescence in AGF strains (46). As such, obtaining a broad representation of the Neocallimastigomycota necessitated the isolation of representatives of various AGF genera *de novo*. Samples were obtained from the feces, rumen, or digesta of domesticated and wild herbivores around the city of Stillwater, OK and Val Verde County, Texas (Table 1). Samples were immediately transferred to the laboratory and the isolation procedures usually commenced within 24 hours of collection. A second round of isolation was occasionally conducted on samples stored at -20^0 C for several weeks (Table 1).

Isolation was performed using a rumen fluid medium reduced by cysteine-sulfide, supplemented with a mixture of kanamycin, penicillin, streptomycin, and chloramphenicol (50 µg/mL, 50 µg/mL, 20 µg/mL, and 50 µg/mL, respectively), and dispensed under a stream of 100% CO_2 (42, 47). All media were prepared according to the Hungate technique (48), as modified by Balch and Wolfe (49). Cellulose (0.5%), or a mixture of switchgrass (0.5%) and cellobiose (0.5%) were used as carbon sources. Samples were serially diluted and incubated at 39°C for 24-48 h. Colonies were obtained from dilutions showing visible signs of fungal growth using the roll tube technique (50). Colonies obtained were inoculated into liquid media, and a second round of isolation and colony picking was conducted to ensure culture purity. Microscopic examination of thallus growth pattern, rhizoid morphology, and zoospore flagellation, as well LSU rRNA gene D1-D2 domain amplification and sequencing were employed to determine the genus level affiliation of all isolates (47). Isolates were maintained and routinely sub-cultured on rumen fluid medium supplemented with antibiotics (to guard

against accidental bacterial contamination) and stored on agar media as described previously (42, 51).

RNA extraction, sequencing, and assembly. Transcriptomic sequencing was conducted for twenty-two AGF strains. Sequencing multiple taxa provides stronger evidence for the occurrence of HGT in a target lineage (52), and allows for the identification of phylum-wide versus genus- and species-specific HGT events. Transcriptomic, rather than genomic, sequencing was chosen for AGF-wide HGT identification efforts since enrichment for polyadenylated (poly(A)) transcripts prior to RNA-seq provides a built-in safeguard against possible prokaryotic contamination, an issue that often plagued eukaryotic genome-based HGT detection efforts (53, 54), as well as to demonstrate that HGT genes identified are transcribed in AGF. Further, sequencing and assembly of a large number of Neocallimastigomycota genomes is challenging due to the extremely high AT content in intergenic regions and the extensive proliferation of microsatellite repeats, often necessitating employing multiple sequencing technologies for successful genomic assembly (42, 45).

Cultures for RNA extraction were grown in rumen fluid medium with cellobiose as the sole carbon source. RNA extraction was conducted on late log/early stationary phase cultures (approximately 48-60 hours post inoculation, depending on strain's growth characteristics) as described previously (55). Briefly, fungal biomass was obtained by vacuum filtration and grounded with a pestle under liquid nitrogen. RNA was extracted using Epicentre MasterPure Yeast RNA Purification kit (Epicentre, Madison, WI, USA) and stored in RNase-free TE buffer. Transcriptomic sequencing using Illumina HiSeq2500 2X150bp paired end technology was conducted using the services of a commercial provider (Novogene Corporation, Beijing, China).

Applied and Environmental Microbiology

RNA-Seq reads were assembled by the de novo transcriptomic assembly program Trinity (56) using previously established protocols (57). All settings were implemented according to the recommended protocol for fungal genomes, with the exception of the absence of the "–jaccard_clip" flag due to the low gene density of anaerobic fungal genomes. The assembly process was conducted on the Oklahoma State University High Performance Computing Cluster as well as the XSEDE HPC Bridges at the Pittsburg Super Computing Center. Quantitative levels for all assembled transcripts were determined using Bowtie2 (58). The program Kallisto was used for quantification and normalization of the gene expression of the transcriptomes (59). All final peptide models predicted were annotated using the Trinotate platform with a combination of homology-based search using BLAST+, domain identification using hmmscan and the Pfam 30.0 database 19 (60), and cellular localization with SignalP 4.0 (61). The twenty-two transcriptomes sequenced in this effort, as well as previously published transcriptomic datasets from *Pecoramyces ruminantium* (42), *Piromyces finnis*, *Piromyces* sp. E2, *Anaeromyces robustus*, and *Neocallimastix californiae* (45) were examined. In each dataset, redundant transcripts were grouped into clusters using CD-HIT-EST with identity parameter of 95% (-c 0.95). The obtained non-redundant transcripts from each analyzed transcriptome were subsequently used for peptide and coding sequence prediction using the TransDecoder with a minimum peptide length of 100 amino acids (http://transdecoder.github.io). Assessment of transcriptome completeness per strain was conducted using BUSCO (62) using Fungi dataset.

HGT identification. A combination of BLAST similarity searches, comparative similarity index (HGT index, h_U), and phylogenetic analyses were conducted to identify HGT events in the analyzed transcriptomic datasets (Figure 1). We define an HGT event as the acquisition of a foreign gene/Pfam by AGF from a single lineage/donor. All predicted peptides were queried

against Uniprot databases (downloaded May 2017) each containing both reviewed (Swiss-Prot) and unreviewed (TrEMBL) sequences. The databases encompassed nine different phylogenetic groups; Bacteria, Archaea, Viridiplantae, Opisthokonta-Chaonoflagellida, Opisthokonta-Fungi (without Neocallimastigomycota representatives), Opisthokonta-Metazoa, Opisthokonta-Nucleariidae and Fonticula group, all other Opisthokonta, and all other non-Opisthokonta-non-Viridiplantae Eukaryota. For each peptide sequence, the bit score threshold and HGT index h_U (calculated as the difference between the bit-scores of the best non-fungal and the best Dikarya fungal matches) were determined. Peptide sequences that satisfied the criteria of having a BLASTP bit-score against a non-fungal database that was >100 (i.e. 2^{-100} chance of random observation) and an HGT index h_U that was ≥30 were considered HGT candidates and subjected to additional phylogenetic analysis. We chose to work with bit-score rather than the raw scores since the bit-score measures sequence similarity independent of query sequence length and database size. This is essential when comparing hits from databases with different sizes (for example, the Bacteria database contained 83 million sequences while the Choanoflagellida database contained 21 thousand sequences). We chose an h_U value of ≥30 (a difference of bit-score of at least 30 between the best non-fungal hit and the best fungal hit to an AGF sequence) previously suggested and validated (63, 64) as the best tradeoff between sensitivity and specificity. Since the bit-score is a logarithmic value that describes sequence similarity, a bit-score > 30 ensure that the sequence aligned much better to the non-fungal hit than it did to the fungal hit.

The identified HGT candidates were modified by removing all CAZyme-encoding sequences (due to their multi-modular nature, see below) and further clustered into orthologues using OrthoMCL (65). Orthologues obtained were subjected to detailed phylogenetic analysis to

confirm HGT occurrence as well as to determine the potential donor. Each Orthologue was queried against the nr database using web Blastp (66) under two different settings: once against the full nr database and once against the Fungi (taxonomy ID: 4751) excluding the Neocallimastigomycetes (Taxonomy ID: 451455). The first 250 hits obtained using these two Blastp searches with an e-value below e^{-10} were downloaded and combined in one fasta file. To remove redundancies, the downloaded sequences were crudely aligned using the standalone Clustal Omega (67) and the alignements were used to generate phylogenetic trees in FastTree under the LG model (68). Produced trees were visualized in FigTree and the groups of sequences that clustered together with very short branches were identified. Perl scripts were then used to remove these redundant sequences from the original fasta files (leaving just one representative). The resulting non-redundant fasta files were used for subsequent analysis. AGF and reference sequences were aligned using MAFFT multiple sequence aligner (69), and alignments were masked for sites with >50% alignment gaps using the Mask Alignment Tool in Geneious 10.2.3 (https://www.geneious.com). Masked alignments were then used in IQ-tree (70) to first predict the best amino acid substitution model (based on the lowest BIC criteria) and to generate maximum likelihood trees under the predicted best model. Both the (-alrt 1000) option for performing the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT), as well as the (-bb 1000) option for ultrafast bootstrap (UFB) (71) were added to the IQ-tree command line. This resulted in the generation of phylogenetic trees with two support values (SH-aLRT and UFB) on each branch. Candidates that showed a nested phylogenetic affiliation that was incongruent to organismal phylogeny with strong SH-aLRT and UFB supports were deemed horizontally transferred. As a final confirmatory step, each tree generated was also reconciled against a species tree (constructed using the large ribosomal subunit L3 protein) using the

programs Ranger-DTL (72) and NOTUNG (73) to infer transfer events at the node where AGF taxa clustered with a phylogenetically-incongruent donor.

Identification of HGT events in carbohydrate active enzymes (CAZymes) transcripts. In AGF genomes, carbohydrate active enzymes (CAZymes) are often encoded by large multi-module genes with multiple adjacent CAZyme or non-CAZyme domains (42, 45). A single gene can hence harbor multiple CAZyme pfams of different (fungal or non-fungal) origins (42, 45). As such, our initial efforts for HGT assessment in CAZyme-encoding transcripts using an entire gene/ transcript strategy yielded inaccurate results since similarity searches only identified pfams with the lowest e-value or highest number of copies, while overlooking additional CAZyme pfams in the transcripts (Figure S2). To circumvent the multi-modular nature of AGF CAZyme transripts, we opted for the identification of CAZyme HGT events on trimmed domains, rather than entire transcript. CAZyme-containing transcripts (Glycoside hydrolases (GHs), Polysaccharide lyases (PLs), and Carbohydrate Esterases (CEs)) were first identified by searching the entire transcriptomic datasets against the dbCAN hidden markov models V5 (74) (downloaded from the dbCAN web server in September 2016) using the command hmmscan in standalone HMMER. For each CAZy family identified, predicted peptides across all transcriptomic datasets were grouped in one fasta file that was then amended with the corresponding Pfam seed sequences (downloaded from the Pfam website (http://pfam.xfam.org/) in March 2017). Sequences were aligned using the standalone Clustal Omega (67) to their corresponding Pfam seeds. Using the Pfam seed sequences as a guide for the start and end of the domain, aligned sequences were then truncated in Jalview (75). Truncated transcripts with an identified CAZy domain were again compared to the pfam database (76) using hmmscan (77) to ensure correct assignment to CAZy families and accurate domain trimming. These truncated

peptide sequences were then analyzed to pinpoint incidents of HGT using the approach described above.

Neocallimastigomycota-specific versus non-specific HGT events. To determine whether an identified HGT event (i.e. foreign gene acquisition from a specific donor) is specific to the phylum Neocallimastigomycota; the occurrence of orthologues (30% identity, >100 amino acids alignment) of the identified HGT genes in basal fungi, i.e. members of Blastocladiales, Chytridiomycota, Cryptomycota, Microsporidia, Mucormycota, and Zoopagomycota, as well as the putative phylogenetic affiliation of these orthologues, when encountered, were assessed. HGT events were judged to be Neocallimastigomycota-specific if: 1. orthologues were absent in all basal fungal genomes, 2. orthologues were identified in basal fungal genomes, but these orthologues were of clear fungal origin, or 3. orthologues were identified in basal fungal genomes and showed a non-fungal phylogenetic affiliation, but such affiliation was different from that observed in the Neocallimastigomycota. On the other hand, events were judged to be non-specific to the Neocallimastigomycota if phylogenetic analysis of basal fungal orthologues indicated a non-fungal origin with a donor affiliation similar to that observed in the Neocallimastigomycota (Figure 1).

Mapping HGT events to available AGF genomes. HGT events identified in AGF datasets examined (both CAZy and non-CAZy events) were mapped onto currently available AGF genome assemblies (42, 45) (Genbank accession numbers ASRE00000000.1, MCOG00000000.1, MCFG00000000.1, MCFH00000000.1). The duplication and expansion patterns, as well as GC content, and intron distribution were assessed in all identified genes. Averages were compared to AGF genome average using Student t-test to identify possible deviations in such characteristics as often observed with HGT genes (78). To avoid any bias the differences in the number of genes

compared might have on the results, we also compared the GC content, codon usage, and intron distribution averages for the identified genes to a subset of an equal number of randomly chosen genes from AGF genomes. We used the MEME Suite's fasta-subsample function (http://meme-suite.org/doc/fasta-subsample.html) to randomly select an equal number of genes from the AGF genomes.

Validation of HGT-identification pipeline using previously published datasets. As a control, the frequency of HGT occurrence in the genomes of a filamentous ascomycete (*Colletotrichum graminicola,* GenBank Assembly accession number GCA_000149035.1), and a microsporidian (*Encephalitozoon hellem,* GenBank Assembly accession number GCA_000277815.3) were determined using our pipeline (Table S1); and the results were compared to previously published results (36, 79).

Guarding against false positive HGT events due to contamination. Multiple safeguards were taken to ensure that the frequency and incidence of HGT reported here are not due to bacterial contamination of AGF transcripts. These included: 1. Application of antibiotics in all culturing procedures as described above. 2. Utilization of transcriptomes rather than genomes selects for eukaryotic polyadenylated (poly(A)) transcripts prior to RNA-seq as a built-in safeguard against possible prokaryotic contamination. 3. Mapping HGT transcripts identified to genomes generated in prior studies and confirming the occurrence of introns in the majority of HGT genes identified. 4. Applying a threshold where only transcripts identified in >50% of transcriptomic assemblies from a specific genus are included and 5. The exclusion of HGT events showing suspiciously high (>90%) sequence identity to donor sequences.

In addition, recent studies have demonstrated that GenBank-deposited reference genomes (53) and transcriptomes (80) of multicellular organisms are often plagued by prokaryotic

297 contamination. The occurrence of prokaryotic contamination in reference donors'
298 genomes/transcriptomes could lead to false positive HGT identification, or incorrect HGT
299 assignments. To guard against any false positive HGT event identification due to possible
300 contamination in reference datasets, sequence data from potential donor reference organisms
301 were queried using blast, and their congruence with organismal phylogeny was considered a
302 prerequisite for inclusion of an HGT event.
303 **Data accession.** Sequences of individual transcripts identified as horizontally transferred are
304 deposited in GenBank under the accession number MH043627-MH043936, and MH044722-
305 MH044724. The whole transcriptome shotgun sequences were deposited in GenBank under the
306 BioProject PRJNA489922, and Biosample accession numbers SAMN09994575-
307 SAMN09994596. Transcriptomic assemblies were deposited in the SRA under project accession
308 number SRP161496. Trees of HGT events discussed in the results and discussion sections are
309 presented in the supplementary document (S5-S45).

Results

Isolates. The transcriptomes of 22 different isolates were sequenced. These isolates belonged to six out of the nine currently described AGF genera: *Anaeromyces* (n=5), *Caecomyces* (n=2), *Neocallimastix* (n=2), *Orpinomyces* (n=3), *Pecoramyces* (n=4), *Piromyces* (n=4), as well as the recently proposed genus *Feramyces* (n=2) (81) (Table 1, Supplementary Fig. 3). Out of the three AGF genera not included in this analysis, two are currently represented by a single strain that was either lost (genus *Oontomyces* (82)), or appears to exhibit an extremely limited geographic and animal host distribution (genus *Buwchfawromyces* (83)). The third unrepresented genus (*Cyllamyces*) has recently been suggested to be phylogenetically synonymous with *Caecomyces* (84). As such, the current collection is a broad representation of currently described AGF genera.

Sequencing. Transcriptomic sequencing yielded 15.2-110.8 million reads (average, 40.87) that were assembled into 31,021-178,809 total transcripts, 17,539-132,141 distinct transcripts (clustering at 95%), and 16,500-70,061 predicted peptides (average 31,611) (Table S2). Assessment of transcriptome completion using BUSCO (62) yielded high values (82.76-97.24%) for all assemblies (Table S1). For strains with a sequenced genome, genome coverage (percentage of genes in a strain's genome for which a transcript was identified) ranged between 70.9-91.4% (Table S2).

HGT events. A total of 12,786 orthologues with a non-fungal bit score > 100, and an HGT index > 30 were identified. After removing orthologues occurring only in a single strain or in less than 50% of isolates belonging to the same genus, 2147 events were further evaluated. Phylogenetic analysis could not confirm the HGT nature (e.g. single long branch that could either be attributed to HGT or gene loss in all other fungi, unstable phylogeny, and/or low bootstrap) of 1863 orthologues and so were subsequently removed. Of the remaining 286

orthologues, 8 had suspiciously high (>90%) first hit amino acid identity. Although the relatively recent divergence and/or acquisition time could explain this high level of similarity, we opted to remove these orthologues as a safeguard against possible bacterial contamination of the transcriptomes. Of the remaining 278 orthologues, one was not inferred as horizontally transferred by the gene-species trees reconciliation softwares used. Ultimately, a total of 277 distinct HGT events that satisfied the criteria described above for HGT were identified (Table S3). The average number of events per genus was 220±12.6 and ranged between 206 in the genus *Orpinomyces* to 237 in the genus *Pecoramyces* pantranscriptomes (Fig. 2A). The majority of HGT acquisition events identified (254, 91.7%) appear to be Neocallimastigomycota-specific, i.e. identified only in genomes belonging to the Neocallimastigomycota, but not in other basal fungal genomes (Table S4), strongly suggesting that such acquisitions occurred post, or concurrent with, the evolution of Neocallimastigomycota as a distinct fungal lineage. As well, the majority of these identified genes were Neocallimastigomycota-wide, being identified in strains belonging to at least six out of the seven examined genera (196 events, 70.76%), suggesting the acquisition of such genes prior to genus level diversification within the Neocallimastigomycota. Only 30 events (10.83%) were genus-specific, with the remainder (51 events, 18.4%) being identified in the transcriptomes of 3-5 genera (Table S4, Figure S4, and Fig. 2b).

The absolute majority (89.2%) of events were successfully mapped to at least one of the four AGF genomes (Table S5), with a fraction (7/30) of the unmapped transcripts being specific to a genus with no genome representative (*Feramyces*, *Caecomyces*). Compared to a random subset of 277 genes in each of the sequenced genomes, horizontally transferred genes in AGF genomes exhibited significantly ($P<0.0001$) fewer introns (1.1±0.31 vs 3.32±0.83), as well as

356 higher GC content (31±4.5 vs 27.7±5.5) (Table S5). Further, HGT genes/pfams often displayed
357 high levels of gene/ pfam duplication and expansion within the genome (Table S5), resulting in
358 an HGT frequency of 2.03% in *Pecoramyces ruminantium* (331 HGT genes out of 16,347 total
359 genes), 2.91% in *Piromyces finnis* (334 HGT genes out of 11,477 total genes), 3.21% in
360 *Anaeromyces robustus* (415 HGT genes out of 12,939 total genes), and 3.46% in *Neocallimastix*
361 *californiae* (724 HGT genes out of 20,939 total genes).
362 **Donors.** A bacterial origin was identified for the majority of HGT events (85.9%), with four
363 bacterial phyla (Firmicutes, Proteobacteria, Bacteroidetes, and Spirochaetes) identified as donors
364 for 169 events (61% of total, 71% of bacterial events) (Fig. 3A). Specifically, the contribution of
365 members of the Firmicutes (119 events) was paramount, the majority of which were most closely
366 affiliated with members of the order Clostridiales (93 events). In addition, minor contributions
367 from a wide range of bacterial phyla were also identified (Fig. 3A). The majority of the putative
368 donor taxa are strict/ facultative anaerobes, and many of which are also known to be major
369 inhabitants of the herbivorous gut and often possess polysaccharide-degradation capabilities (85,
370 86). Archaeal contributions to HGT were extremely rare (5 events). On the other hand, multiple
371 (30) events with eukaryotic donors were identified. In few instances, a clear non-fungal origin
372 was identified for a specific event, but the precise inference of the donor based on phylogenetic
373 analysis was not feasible (Table S4).
374 **Metabolic characterization.** Functional annotation of HGT genes/pfams indicated that the
375 majority (63.9%) of events encode metabolic functions such as extracellular polysaccharide
376 degradation and central metabolic processes. Bacterial donors were slightly overrepresented in
377 metabolic HGT events (87.5% of the metabolism-related events, compared to 85.9% of the total
378 events). Genes involved in cellular processes and signaling represent the second most

represented HGT events (11.19%), while genes involved in information storage and processing only made up 4.69% of the HGT events identified (Figs 3b-e). Below we present a detailed description of the putative abilities and functions enabled by HGT transfer events.

Central catabolic abilities. Multiple HGT events encoding various central catabolic processes were identified in AGF transcriptomes and successfully mapped to the genomes (Fig. 4, Table S4, Figs S5-S16). A group of events appears to encode enzymes that allow AGF to channel specific substrates into central metabolic pathways. For example, genes encoding enzymes of the Leloir pathway for galactose conversion to glucose-1-phosphate (galactose-1-epimerase, galactokinase (Fig. 5A), and galactose-1-phosphate uridylyltransferase) were identified, in addition to genes encoding ribokinase, as well as xylose isomerase and xylulokinase for ribose and xylose channeling into the pentose phosphate pathway. As well, genes encoding deoxyribose-phosphate aldolase (DeoC) enabling the utilization of purines as carbon and energy sources were also horizontally acquired in AGF. Further, several of the glycolysis/gluconeogenesis genes, e.g. phosphoenolpyruvate synthase, as well as phosphoglycerate mutase were also of bacterial origin. Fungal homologs of these glycolysis/gluconeogenesis genes were not identified in the AGF transcriptomes and genomes, suggesting the occurrence of xenologous replacement HGT events. In addition to broadening substrate range, HGT acquisitions provided additional venues for recycling reduced electron carriers via new fermentative pathways in this strictly anaerobic and fermentative lineage. The production of ethanol, D-lactate, and hydrogen appears to be enabled by HGT (Fig. 4). The acquisition of several aldehyde/alcohol dehydrogenases, and of D-Lactate dehydrogenase for ethanol and lactate production from pyruvate was identified. Although these two enzymes are encoded in other fungi as part of their fermentative capacity (e.g. *Saccharomyces* and *Schizosaccharomyces*), no homologs of these fungal genes were identified in

AGF pantranscriptomes. Hydrogen production in AGF, as well as in many anaerobic eukaryotes with mitochondria-related organelles (e.g. hydrogenosomes and mitosomes), involves pyruvate decarboxylation to acetyl CoA, followed by the use of electrons generated for hydrogen formation via an anaerobic Fe-Fe hydrogenase (42, 87, 88). In AGF, while enzymes for pyruvate decarboxylation to acetyl CoA (pyruvate-formate lyase) and the subsequent production of acetate in the hydrogenosome (via acetyl-CoA:succinyl transferase) appear to be of fungal origin, the Fe-Fe hydrogenase and its entire maturation machinery (HydEFG) seem to be horizontally transferred being phylogenetically affiliated with similar enzymes in Thermotogae, Clostridiales, and the anaerobic jakobid excavate, *Stygiella incarcerate* (Fig. 5B). It has recently been suggested that *Stygiella* acquired the Fe-Fe hydrogenase and its maturation machinery from bacterial donors including Thermotogae, Firmicutes, and Spirochaetes (89), suggesting either a single early acquisition event in eukaryotes, or alternatively independent events for the same group of genes have occurred in different eukaryotes. With the exception of the Fe-Fe hydrogenase and its maturation machinery, no other hydrogenosomally-destined proteins (see list in reference (42)) were identified as horizontally transferred in this study. These results collectively suggest that HGT did not play a role in the evolution of hydrogenosomes in AGF; and reinforces the proposed mitochondrial origin of hydrogenosomes through reductive evolution (88).

Anabolic capabilities. Multiple anabolic genes that expanded AGF biosynthetic capacities appear to be horizontally transferred (Fig. S17-S30). These include several amino acid biosynthesis genes e.g. cysteine biosynthesis from serine; glycine and threonine interconversion; and asparagine synthesis from aspartate. In addition, horizontal gene transfer allowed AGF to de-novo synthesize NAD via the bacterial pathway (starting from aspartate via L-aspartate oxidase

(NadB; Fig. 5C) and quinolinate synthase (NadA) rather than the five-enzymes fungal pathway starting from tryptophan (90)). HGT also allowed AGF to salvage thiamine via the acquisition of phosphomethylpyrimidine kinase. Additionally, several genes encoding enzymes in purine and pyrimidine biosynthesis were horizontally transferred (Fig. 4). Finally, horizontal gene transfer allowed AGF to synthesize phosphatidyl-serine from CDP-diacylglycerol, and to convert phosphatidyl-ethanolamine to phosphatidyl-choline.

Adaptation to the host environment. Horizontal gene transfer also appears to have provided means of guarding against toxic levels of compounds known to occur in the host animal gut (Fig. S31-S37). For example, methylglyoxal, a reactive electrophilic species (91), is inevitably produced by ruminal bacteria from dihydroxyacetone phosphate when experiencing growth conditions with excess sugar and limiting nitrogen (92). Genes encoding enzymes mediating methylglyoxal conversion to D-lactate (glyoxalase I and glyoxalase II-encoding genes) appear to be acquired via HGT in AGF. Further, HGT allowed several means of adaptation to anaerobiosis. These include: 1) acquisition of the oxygen-sensitive ribonucleoside-triphosphate reductase class III (Fig. 5D) that is known to only function during anaerobiosis to convert ribonucleotides to deoxyribonucleotides (93), 2) acquisition of squalene-hopene cyclase, which catalyzes the cyclization of squalene into hopene, an essential step in biosynthesis of the cell membrane steroid tetrahymanol that replaced the molecular O_2-requiring ergosterol in the cell membranes of AGF, 3) acquisition of several enzymes in the oxidative stress machinery including Fe/Mn superoxide dismutase, glutathione peroxidase, rubredoxin/rubrerythrin, and alkylhydroperoxidase.

In addition to anaerobiosis, multiple horizontally transferred general stress and repair enzymes were identified (Fig. S38-S45). HGT-acquired genes encoding 2-phosphoglycolate

phosphatase, known to metabolize the 2-phosphoglycolate produced in the repair of DNA lesions induced by oxidative stress (94) to glycolate, were identified in all AGF transcriptomes studied (Fig. 4, Table S4). Surprisingly, two genes encoding antibiotic resistance enzymes, chloramphenicol acetyltransferase and aminoglycoside phosphotransferase, were identified in all AGF transcriptomes, presumably to improve its fitness in the eutrophic rumen habitat that harbors antibiotic-producing prokaryotes (Table S4). While unusual for eukaryotes to express antibiotic resistance genes, basal fungi such as *Allomyces*, *Batrachochytrium*, and *Blastocladiella* were shown to be susceptible to chloramphenicol and streptomycin (95, 96). Other horizontally transferred repair enzymes include DNA-3-methyladenine glycosylase I, methylated-DNA--protein-cysteine methyltransferase, galactoside and maltose O-acetyltransferase, and methionine-R-sulfoxide reductase (Table S4).

HGT transfer in AGF carbohydrate active enzymes machinery. Within the analyzed AGF transcriptomes, CAZymes belonging to 39 glycoside hydrolase (GHs), 5 polysaccharide lyase (PLs), and 10 carbohydrate esterase (CEs) families were identified (Fig. 6). The composition of the CAZymes of various AGF strains examined were broadly similar, with the following ten notable exceptions: Presence of GH24 and GH78 transcripts only in *Anaeromyces* and *Orpinomyces*, the presence of GH28 transcripts only in *Pecoramyces*, *Neocallimastix*, and *Orpinomyces*, the presence of GH30 transcripts only in *Anaeromyces*, and *Neocallimastix*, the presence of GH36 and GH95 transcripts only in *Anaeromyces*, *Neocallimastix*, and *Orpinomyces*, the presence of GH97 transcripts only in *Neocallimastix*, and *Feramyces*, the presence of GH108 transcripts only in *Neocallimastix*, and *Piromyces*, and the presence of GH37 predominantly in *Neocallimastix*, GH57 transcripts predominantly in *Orpinomyces*, GH76 transcripts predominantly in *Feramyces*, and CE7 transcripts predominantly in *Anaeromyces* (Fig. 6).

471 HGT appears to be rampant in the AGF pan-CAZyome: A total of 72 events (26% of total HGT
472 events) were identified, with 40.3% occurring in at least 6 of the 7 AGF genera examined (Fig. 6,
473 Table S4). In 48.7% of GH families, 50% of CE families, and 40% of PL families, a single event
474 (i.e. attributed to one donor) was observed (Fig. 6, Table S4).
475 Duplication of these events in AGF genomes was notable, with 132, 310, 156, and 130 copies of
476 HGT CAZyme pfams identified in *Anaeromyces*, *Neocallimastix*, *Piromyces* and *Pecoramyces*
477 genomes, representing 33.59%, 36.77%, 40.41%, and 24.62% of the overall organismal
478 CAZyme machinery (Table S5). The contribution of Viridiplantae, Fibrobacteres, and Gamma-
479 Proteobacteria was either exclusive to CAZyme-related HGT events or significantly higher in
480 CAZyme, compared to other events (Fig. 3A).
481 Transcripts acquired by HGT represented >50% of transcripts in anywhere between 13
482 (*Caecomyces*) to 20 (*Anaeromyces*) GH families; 3 (*Caecomyces*) to 5 (*Anaeromyces*,
483 *Neocallimastix*, *Orpinomyces*, and *Feramyces*) CE families; and 2 (*Caecomyces* and *Feramyces*)
484 to 3 (*Anaeromyces*, *Pecoramyces*, *Piromyces*, *Neocallimastix*, and *Orpinomyces*) PL families
485 (Fig. 6). It is important to note that in all these families, multiple transcripts appeared to be of
486 bacterial origin based on BLAST similarity search but did not meet the strict criteria
487 implemented for HGT determination in this study. As such, the contribution of HGT transcripts
488 to overall transcripts in these families is probably an underestimate. Only GH9, GH20, GH37,
489 GH45, and PL3 families appear to lack any detectable HGT events. A PCA biplot comparing
490 CAZyomes in AGF genomes to other basal fungal lineages strongly suggests that the acquisition
491 and expansion of many of these foreign genes play an important role in shaping the
492 lignocellulolytic machinery of AGF (Fig. 7). The majority of CAZyme families defining AGF
493 CAZyome were predominantly of non-fungal origin (Fig. 7). This pattern clearly attests to the

value of HGT in shaping AGF CAZyome via acquisition and extensive duplication of acquired gene families.

Collectively, HGT had a profound impact on AGF plant biomass degradation capabilities, as recently proposed (97). The AGF CAZyome encodes enzymes putatively mediating the degradation of twelve different polysaccharides (Fig. S46). In all instances, GH and PL families with >50% horizontally transferred transcripts contributed to backbone cleavage of these polymers; although in many polymers, e.g. cellulose, glucoarabinoxylan, and rhamnogalactouronan, multiple different GHs can contribute to backbone cleavage. Similarly, GH, CE, and PL families with >50% horizontally transferred transcripts contributed to 10 out of 13 side-chain-cleaving activities, and 3 out of 5 oligomer-to-monomer breakdown activities (Fig. S46).

Discussion

Here, we present a systematic analysis of HGT patterns in 27 transcriptomes and 4 genomes belonging to the Neocallimastigomycota. Our analysis identified 277 events, representing 2-3.46% of genes in examined AGF genomes. Further, we consider these values to be conservative estimates due to the highly stringent criteria and employed. Only events with h_U of >30 were considered, and all putative events were further subjected to manual inspection, phylogenetic tree construction, and gene-species tree reconciliation analysis to confirm incongruence with organismal evolution and bootstrap-supported affiliation to donor lineages. Further, events identified in less than 50% of strains in a specific genus were excluded, and parametric gene composition approaches were implemented in conjunction with sequence-based analysis.

Multiple factors could be postulated to account for the observed high HGT frequency in AGF. The sequestration of AGF into the anaerobic, prokaryotes-dominated herbivorous gut necessitated the implementation of the relatively faster adaptive mechanisms for survival in this new environment, as opposed to the slower strategies of neofunctionalization and gene birth. Indeed, niche adaptation and habitat diversification events are widely considered important drivers for HGT in eukaryotes (16, 23, 26, 98)(37). Further, AGF are constantly exposed to a rich milieu of cells and degraded DNA in the herbivorous gut. Such close physical proximity between donors/ extracellular DNA and recipients is also known to greatly facilitate HGT (99-101). Finally, AGF release asexual motile free zoospores into the herbivorous gut as part of their life cycle (40). According to the weak-link model (102), these weakly protected and exposed structures provide excellent entry point of foreign DNA to eukaryotic genomes. It is important to note that AGF zoospores also appear to be naturally competent, capable of readily uptaking nucleic acids from their surrounding environment (51).

528 The anaerobic gut fungi have a notoriously low GC content, ranging between 13-20%. It
529 has previously been postulated that this low GC content is due to genetic drift (42) triggered by
530 the low effective population sizes, bottlenecks in vertical transmission, and the asexual life style
531 of anaerobic fungi. As such, the low GC content is an additional consequence of AGF
532 sequestration in the herbivorous gut. Whether the low GC content in AGF played a role in
533 facilitating HGT is currently unclear. It is worth mentioning, however, that the majority of AGF
534 donors identified in this study are members of the bacterial order Clostridiales, many of which
535 have relatively low GC content genomes.
536 The distribution of HGT events across various AGF taxa (Fig. 2), identities of HGT donors (Fig.
537 3), and abilities imparted (Figs. 4-5) could offer important clues regarding the timing and impact
538 of HGT on Neocallimastigomycota evolution. The majority of events (70.76%) were
539 Neocallimastigomycota-wide and were mostly acquired from lineages known to inhabit the
540 herbivorous gut, e.g. Firmicutes, Proteobacteria, Bacteroidetes, and Spirochaetes (Figs. 2-3).
541 This pattern strongly suggests that such acquisitions occurred post (or concurrent with) AGF
542 sequestration into the herbivorous gut, but prior to AGF genus level diversification. Many of the
543 functions encoded by these events represented novel functional acquisitions that impart new
544 abilities, e.g. galactose metabolism, methyl glyoxal detoxification, pyruvate fermentation to d-
545 lactate and ethanol, and chloramphenicol resistance (Fig. 3). Others represented acquisition of
546 novel genes or pfams augmenting existing capabilities within the AGF genomes, e.g. acquisition
547 of GH5 cellulases to augment the fungal GH45, acquisition of additional GH1 and GH3 beta
548 gluco- and galactosidases to augment similar enzymes of apparent fungal origin in AGF
549 genomes (Fig. 6-7, Fig. S46). Novel functional acquisition events enabled AGF to survive and
550 colonize the herbivorous gut by: 1. Expanding substrate-degradation capabilities (Fig. 5a, 6, 7,

551 S5-S17, Table S4), hence improving fitness by maximizing carbon and energy acquisition from
552 available plant substrates, 2. Providing additional venues for electron disposal via lactate, ethanol,
553 and hydrogen production, and 3. Enabling adaptation to anaerobiosis (Fig. 4, S32-S38, Table S4).
554 A smaller number of observed events (n=30) were genus-specific (Fig. 2, Table S4). This
555 group was characterized by being significantly enriched in CAZymes (56.7% of genus-specific
556 horizontally transferred events have a predicted CAZyme function, as opposed to 26% in the
557 overall HGT dataset), and being almost exclusively acquired from donors that are known to
558 inhabit the herbivorous gut (103) (25 out of the 30 events were acquired from the orders
559 Clostridiales, Bacillales, and Lactobacillales within Firmicutes, Burkholderiales within the Beta-
560 Proteobacteria, Flavobacteriales and Bacteroidales within Bacteroidetes, and the Spirochaetes,
561 Actinobacteria, and Lentisphaerae), or from Viridiplantae (4 out of the 30 events). Such pattern
562 suggests the occurrence of these events relatively recently in the herbivorous gut post AGF genus
563 level diversification. A recent study also highlighted the role of HGT in complementing the
564 CAZyme machinery of *Piromyces* sp. strain E2 (97). We reason that the lower frequency of such
565 events is a reflection of the relaxed pressure for acquisition and retention of foreign genes at this
566 stage of AGF evolution.
567 Gene acquisition by HGT necessitates physical contact between donor and recipient
568 organisms. Many of the HGT acquired traits by AGF are acquired from prokaryotes that are
569 prevalent in the herbivorous gut microbiota (Fig. 3). However, since many of these traits are
570 absolutely necessary for survival in the gut, the establishment of AGF ancestors in this
571 seemingly inhospitable habitat is, theoretically, unfeasible. This dilemma is common to all HGT
572 processes enabling niche adaptation and habitat diversification (22). We put forth two
573 evolutionary scenarios that could explain this dilemma not only for AGF, but also for other gut-

574 dwelling anaerobic microeukaryotes, e.g. *Giardia*, *Blastocystis*, and *Entamoeba*, where HGT was
575 shown to play a vital role in enabling survival in anaerobic conditions (22, 104, 105). The first is
576 a coevolution scenario in which the progressive evolution of the mammalian gut from a short and
577 predominantly aerobic structure characteristic of carnivores/insectivores to the longer, more
578 complex, and compartmentalized structure encountered in herbivores was associated with a
579 parallel progressive and stepwise acquisition of genes required for plant polymers metabolism
580 and anaerobiosis by AGF ancestors, hence assuring its survival and establishment in the current
581 herbivorous gut. The second possibility is that AGF ancestors were indeed acquired into a
582 complex and anaerobic herbivorous gut, but initially represented an extremely minor component
583 of the gut microbiome and survived in locations with relatively higher oxygen concentration in
584 the alimentary tract e.g. mouth, saliva, esophagus or in micro-niches in the rumen where
585 transient oxygen exposure occurs. Subsequently, HGT acquisition has enabled the expansion of
586 their niche, improved their competitiveness and their relative abundance in the herbivorous gut to
587 the current levels.

588 In conclusion, our survey of HGT in AGF acquisition demonstrates that the process is
589 absolutely crucial for the survival and growth of AGF in its unique habitat. This is not only
590 reflected in the large number of events, massive duplication of acquired genes, and overall high
591 HGT frequency observed in AGF genomes, but also in the nature of abilities imparted by the
592 process. HGT events not only facilitated AGF adaptation to anaerobiosis, but also allowed them
593 to drastically improve their polysaccharide degradation capacities, provide new venues for
594 electron disposal via fermentation, and acquire new biosynthetic abilities. As such, we reason
595 that the process should not merely be regarded as a conduit for supplemental acquisition of few
596 additional beneficial traits. Rather, we posit that HGT enabled AGF to forge a new evolutionary

trajectory, resulting in Neocallimastigomycota sequestration, evolution as a distinct fungal lineage in the fungal tree of life, and subsequent genus and species level diversification. This provides an excellent example of the role of HGT in forging the formation of high rank taxonomic lineages during eukaryotic evolution.

Conflict of Interest. The authors declare no conflict of interest.

Acknowledgments. This work has been funded by the NSF-DEB Grant numbers 1557102 to N.Y. and M.E. and 1557110 to J.E.S.

References:

1. Doolittle WF. 1999. Lateral Genomics. Trends Cell Biol 9:M5-M8.
2. Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97-10.
3. Cai J, Zhao R, Jiang H, Wang W. 2008. De Novo origination of a new protein-coding gene in *Saccharomyces cerevisiae*. Genetics 179:487-496.
4. Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res 20:1313-1326.
5. Andersson DI, Jerlström-Hultqvist J, Näsvall J. 2015. Evolution of new functions de novo and from preexisting genes. Cold Spring Harb Perspect Biol 7:a017996.
6. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B, Brar GA, Weissman JS, Regev A, Thierry-Mieg N, Cusick ME, Vidal M. 2010. Proto-genes and de novo gene birth. Nature 487:370-374.
7. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299-304.
8. Caro-Quintero A, Konstantinidis K. 2015. Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. ISME J 9:958-967.
9. Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS. 2015. Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. ISME J 9:447-460.
10. Puigbo P, Pasamontes A, Garcia-Vallve S. 2008. Gaining and losing the thermophilic adaptation in prokaryotes. Trends Genet 24:10-14.
11. Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Daly MJ, Koonin EV, Makarova KS. 2005. Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. BMC Evol Biol 5:57.
12. Wiedenbeck J, Cohan FM. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev 35:957-976.
13. Syvanen M. 2012. Evolutionary implications of horizontal gene transfer. Annu Rev Genet 46:341-358.
14. Philippe H, Douady CJ. 2003. Horizontal gene transfer and phylogenetics. Curr Opin Microbiol 6:498-505.
15. Fitzpatrick DA. 2012. Horizontal gene transfer in fungi. FEMS Microbiol Lett 2011:1-8.
16. Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605-618.
17. Husnik F, McCutcheon JP. 2017. Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol 16:67–79.
18. Qian Q, Keeling PJ. 2001. Diplonemid glyceraldehyde–3-phosphate dehydrogenase (GAPDH) and prokaryoteto-eukaryote lateral gene transfer. Protist 152:193–201.
19. Hirt RP, Harriman N, Kajava AV, Embley TM. 2002. A novel potential surface protein in Trichomonas vaginalis contains a leucine-rich repeat shared by micro-organisms from all three domains of life. Mol Biochem Parasitol 125:195–199.
20. Nixon JEJ, Wang A, Field J, Morrison HG, McArthur AG, Sogin ML, al. e. 2002. Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH

oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. Eukaryot Cell 1:181–190.

21. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, al. e. 2005. The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43-57.
22. Eme L, Gentekaki E, Curtis B, Archibald JM, Roger AJ. 2017. Lateral gene transfer in the adaptation of the anaerobic parasite *Blastocystis* to the gut. Curr Biol 27:807-820.
23. Ricard G, McEwan NR, Dutilh BE, Jouany J-P, Macheboeuf D, Mitsumori M, McIntosh FM, Michalowski T, Nagamine T, Nelson N, Newbold CJ, Nsabimana E, A.Takenaka, Thomas NA, Ushida K, Hackstein JH, Huynen MA. 2006. Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. BMC Genomics 7:22.
24. Kishore SP, Stiller JW, Deitsch KW. 2013. Horizontal gene transfer of epigenetic machinery and evolution of parasitism in the malaria parasite *Plasmodium falciparum* and other apicomplexans. BMC Evol Biol 13:37.
25. Wisecaver JH, Brosnahan ML, Hackett JD. 2013. Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. Genome Biol Evol 5:2368-2381.
26. Schönknecht G, Weber AP, Lercher MJ. 2013. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. Bioassays 36:9-20.
27. Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. J Exp Bot 58:1-9.
28. Sun GL, Yang ZF, Ishwar A, Huang JL. 2010. Algal genes in the closest relatives of animals. Mol Biol Evol 27:2879-2889.
29. Marcet-Bouben M, Gabaldon T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet 26:5-8.
30. Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. Science 320:1210-1213.
31. Danchin EG, Rosso MN, Vieira P, Almeida-Engler Jd, Coutinho PM, Henrissat B, Abad P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Nat Acad Sci USA 107:17651-17656.
32. McCarthy CGP, Fitzpatrick DA. 2016. Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to oomycete lineages. mSphere 1:e00195-16.
33. Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc Nat Acad Sci USA, 101:15386-15391.
34. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Nat Acad Sci USA 108:13624-13629.
35. Taylor JW, Berbee ML. 2006. Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia 98:838-49.
36. Alexander WG, Wisecaver JH, Rokas A, Httinger CT. 2016. Horizontally aquired gene in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. Proc Nat Acad Sci USA 113:4116-4121.
37. Druzhinina IS, Chenthamara K, Zhang J, Atanasova L, Yang D, Miao Y, Rahimi MJ, Grujic M, Cai F, Pourmehdi S, Salim KA, Pretzer C, Kopchinskiy AG, Henrissat B, Kuo A, Hundley H, Wang M, Aerts A, Salamov A, Lipzen A, LaButti K, Barry K, Grigoriev IV, Shen Q, Kubicek CP. 2018. Massive lateral transfer of genes encoding plant cell

wall-degrading enzymes to the mycoparasitic fungus *Trichoderma* from its plant-associated hosts. PLoS Genet 14:e1007322.

38. Berbee ML, James TY, Strullu-Derrien C. 2017. Early diverging fungi: diversity and impact at the dawn of terrestrial life. Annu Rev Microbiol 71:41-60.

39. Doolittle WF. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307–311.

40. Gruninger RJ, Puniyab AK, Callaghanc TM, Edwardsc JE, Youssef N, Dagare SS, Fliegerova K, Griffith GW, Forster R, Tsang A, McAllister T, Elshahed MS. 2014. Anaerobic Fungi (Phylum Neocallimastigomycota): Advances in understanding of their taxonomy, life cycle, ecology, role, and biotechnological potential. FEMS Microbiol Ecol 90:1-17.

41. Boxma B, Voncken F, Jannink S, Alen TV, Akhmanova A, Weelden SWHV, Hellemond JJV, Ricard G, Huynen M, Tielens AGM, Hackstein JHP. 2004. The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. Mol Microbiol 51:1389 -1399.

42. Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ, Atiyeh HK, Wilkins MR, Elshahed MS. 2013. Genome of the anaerobic fungus *Orpinomyces* sp. C1A reveals the unique evolutionary history of a remarkable plant biomass degrader Appl Environ Microbiol 79:4620-4634.

43. Harhangi HR, Akhmanova AS, Emmens R, Drift Cvd, Laat WTAMd, Dijken JPv, Jetten MSM, Pronk JT, Camp HJMOd. 2003. Xylose metabolism in the anaerobic fungus *Piromyces* sp. strain E2 follows the bacterial pathway. Arch Microbiol 180:134-142.

44. Garcia-Vallvé S, Romeu A, Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen Fungi. Mol Biol Evol 17:352-361.

45. Haitjema CH, Gilmore SP, Henske JK, Solomon KV, Groot Rd, Kuo A, Mondo SJ, Salamov AA, LaButti K, Zhao Z, Chiniquy J, Barry K, Brewer HM, Purvine SO, Wright AT, Hainaut M, Boxma B, Alen Tv, Hackstein JHP, Henrissat B, Baker SE, Grigoriev IV, O'Malley MA. 2017. A parts list for fungal cellulosomes revealed by comparative genomics. Nature Microbiol 2:17087.

46. Ho YW, Barr DJS. 1995. Classification of anaerobic gut fungi from herbivores with emphasis on rumen fungi from malaysia. Mycologia 87:655-677.

47. Hanafy RA, Elshahed MS, Liggenstoffer AS, Griffith GW, Youssef NH. 2017. *Pecoramyces ruminantium*, gen. nov., sp. nov., an anaerobic gut fungus from the feces of cattle and sheep. Mycologia 109:231-243.

48. Bryant M. 1972. Commentary on the Hungate technique for culture of anaerobic bacteria. Am J Clin Nutr 25:1324-1328.

49. Balch WE, Wolfe R. 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressureized atmosphere. Appl Environ Microbiol 32:781-791.

50. Hungate RE. 1969. A roll tube method for cultivation of strict anaerobes. Meth Microbiol 3:117-132.

51. Calkins S, Elledge NC, Hanafy RA, Elshahed MS, Youssef NH. 2016. A fast and reliable procedure for spore collection from anaerobic fungi: Application for RNA uptake and long-term storage of isolates. J Microbiol Methods 127:206-213.

52. Richards TA, Monier A. 2017. A tale of two tradigrades. Proc Nat Acad Sci USA 113:4892-4894.

53. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, Tintori SC, Li Q, Jones CD, Yandell M, Messina DN, Glasscock J, Goldstein B. 2015. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proc Nat Acad Sci USA 112:15976-15981.

54. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M. 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. Proc Nat Acad Sci USA 113:5053-5058.

55. Couger MB, Youssef NH, Struchtemeyer CG, Liggenstoffer AS, Elshahed MS. 2015. Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate *Orpinomyces* sp. strain C1A. Bitechnol Biofuels 8:208.

56. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma Fd, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.

57. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger M, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols 8:1494-1512.

58. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

59. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525-527.

60. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. 44:D279-D285.

61. Petersen TN, Brunak S, Heijne Gv, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786.

62. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.

63. Boschetti C, Carr A, Crisp A, Eyres I, Wang-Koh Y, Lubzens E, Barraclough TG, Micklem G, Tunnacliffe A. 2012. Biochemical diversification through foreign gene expression in bdelloid rotifers. PLOS Genet 8:e1003035.

64. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. 2015. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol 16:50.

65. Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178-2189.

66. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

67. Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 27:135-145.

68. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.
69. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772-780.
70. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268-274.
71. Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 30:1188-95.
72. Kordi M, Kundu S, Bansal MS, Kellis M. 2018. RANGER-DTL 2.0: rigorous reconstruction of gene-family evolution by duplication, transfer and loss. Bioinformatics 34:3214-3216.
73. Lai H, Stolzer M, Durand D, Xu M, Sathaye D, Vernot B. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics 28:i409-i415.
74. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445-W451.
75. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189-91.
76. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222-30.
77. Prakash A, Jeffryes M, Bateman A, Finn RD. 2017. The HMMER Web Server for Protein Sequence Similarity Search. Curr Protoc Bioinformatics 60:3.15.1-3.15.23.
78. Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. Nat Rev Genet 16:472-82.
79. Jaramillo VDA, Sukno SA, Thon MR. 2015. Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer. BMC Genomics 16:2.
80. Stairs CW, Eme L, Muñoz-Gómez SA, Cohen A, Dellaire G, Shepherd JN, Fawcett JP, Roger AJ. 2018. Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. eLife 7:e34292.
81. Hanafy RA, Elshahed MS, Youssef NH. 2018. *Feramyces austinii*, gen. nov, sp. nov., an anaerobic gut fungus from rumen and fecal samples of wild Barbary sheep and fallow deer Submitted.
82. Dagar SS, Kumar S, Griffith GW, Edwards JE, Callaghan TM, Singh R, Nagpal AK, Puniya AK. 2015. A new anaerobic fungus (*Oontomyces anksri* gen. nov., sp. nov.) from the digestive tract of the Indian camel (*Camelus dromedarius*). Fungal Biol 19:731-737.
83. Callaghan TM, Podmirseg SM, Hohlweck D, Edwards JE, Puniya AK, Dagar SS, Griffith GW. 2015. *Buwchfawromyces eastonii* gen. nov., sp. nov.: a new anaerobic fungus (Neocallimastigomycota) isolated from buffalo faeces. Mycokeys 9:11-28.
842. Wang X, Liu X, Groenewald JZ. 2017. Phylogeny of anaerobic fungi (phylum Neocallimastigomycota), with contributions from yak in China. Antonie Van Leeuwenhoek 110:87-103.

85. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nature Commun 9:870.
86. He J, Yi L, Hai L, Ming L, Gao W, Ji R. 2018. Characterizing the bacterial microbiota in different gastrointestinal tract segments of the Bactrian camel. Sci Rep 8:654.
87. Müller M, Mentel M, Hellemond JJv, Henze K, Woehle C, Gould SB, Yu R-Y, Giezen Mvd, Tielens AGM, Martin WF. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol Mol Biol Rev 76:444–495.
88. Yarlett N, Hackstein JHP. 2005. Hydrogenosomes: one organelle, multiple origins BioScience 55:657–668.
89. Leger MM, Eme L, Hug LA, Roger AJ. 2016. Novel hydrogenosomes in the microaerophilic jakobid *Stygiella incarcerata*. Mol Biol Evol 33:2318-2336.
90. Lin H, Kwan AL, Dutcher SK. 2010. Synthesizing and Salvaging NAD+: Lessons Learned from *Chlamydomonas reinhardtii*. PLOS Genetics 6:e1001105.
91. Lee C, Park C. 2017. Bacterial responses to glyoxal and methylglyoxal: reactive electrophilic species. Int J Mol Sci 18:169.
92. Russell JB. 1993. Glucose toxicity in *Prevotella ruminicola*: methylglyoxal accumulation and its effect on membrane physiology. Appl Environ Microbiol 59:2844-2850.
93. Jordan A, Reichard P. 1998. Ribonucleotide reductases. Ann Rev Biochem 67:71-98.
94. Pellicer MT, Nuñez MF, Aguilar J, Badia J, Baldoma L. 2003. Role of 2-phosphoglycolate phosphatase of *Escherichia coli* in metabolism of the 2-phosphoglycolate formed in DNA repair. J Bacteriol 185:5815-5821.
95. Rooke DM, Shattock RC. 1983. Effect of chloramphenicol and streptomycin on developmental stages of *Phytophthora infestans*. Microbiology 129:3401-3410.
96. Bishop PJ, Speare R, Poulter R, Butler M, Speare BJ, Hyatt A, Olsen V, Haigh A. 2009. Elimination of the amphibian chytrid fungus *Batrachochytrium dendrobatidis* by Archey's frog *Leiopelma archeyi*. Dis Aquat Organ 84:9-15.
97. Duarte I, Huynen MA. 2019. Contribution of lateral gene transfer to the evolution of the eukaryotic fungus *Piromyces* sp. E2: Massive bacterial transfer of genes involved in carbohydrate metabolism. BioRxiv:514042.
98. de Koning AP, Brinkman FS, Jones SJ, Keeling PJ. 2000. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. Mol Biol Evol 17:1769-1773.
99. Shterzer N, Mizrahi I. 2015. The animal gut as a melting pot for horizontal gene transfer. Can J Microbiol 61:603-605.
100. Moliner C, Fournier PE, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol Rev 34:281-294
101. Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc Nat Acad Sci USA 102:14332-14337.
102. Huang JL. 2013. Horizontal gene transfer in eukaryotes: the weak-link model. Bioassays 35:868-875.
103. Creevey CJ, Kelly WJ, Henderson G, Leahy SC. 2014. Determining the culturability of the rumen bacterial microbiome. Microb Biotechnol 7:467-479.

874 104. Andersson JO, Sjögren AM, Davis LA, Embley TM, Roger AJ. 2003. Phylogenetic
875 analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes.
876 Curr Biol 13:94-104.
877
878 105. Grant JR, Katz LA. 2014. Phylogenomic study indicates widespread lateral gene transfer
879 in Entamoeba and suggests a past intimate relationship with parabasalids. Genome Biol
880 Evol 6:2350-2360.
881 106. Teunissen MJ, Op den Camp HJ, Orpin CG, Huis in 't Veld JH, Vogels GD. 1991.
882 Comparison of growth characteristics of anaerobic fungi isolated from ruminant and non-
883 ruminant herbivores during cultivation in a defined medium. J Gen Microbiol 137:1401-8.
884 107. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK,
885 Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N,
886 Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017.
887 CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.
888 Nucleic Acids Res 45:D200-D203.
889 108. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and
890 interpretation of large-scale molecular data sets. Nucleic Acids Res 40:D109-14.
891

892 Figure Legends

893 **Figure 1**. Workflow diagram describing the procedure employed for identification HGT events
894 in Neocallimastigomycota datasets analyzed in this study.

895 **Figure 2.** (A) Distribution pattern of HGT events in AGF transcriptomes demonstrating that the
896 majority of events were Neocallimastigomycota-wide i.e. identified in all seven AGF genera
897 examined. (B) Total Number of HGT events identified per AGF genus.

898 **Figure 3.** Identity of HGT donors and their contribution to the various functional classes. The X-
899 axis shows the absolute number of events belonging to each of the functional classes shown in
900 the legend. The tree is intended to show the relationship between the donors' taxa and is not
901 drawn to scale. Bacterial donors are shown with red branches depicting the phylum-level, with
902 the exception of Firmicutes and Bacteroidetes donors, where the order-level is shown, and
903 Proteobacteria, where the class-level is shown. Archaeal donors are shown with green branches
904 and all belonged to the Methanobacteriales order of Euryarchaeota. Eukaryotic donors are shown
905 with blue branches. Only the 230 events from a definitive-taxon donor are shown in the figure.
906 The other 53 events were clearly nested within a non-fungal clade, but a definitive donor taxon
907 could not be ascertained. Functional classification of the HGT events, determined by searching
908 the Conserved Domain server (107) against the COG database are shown in B. For events with
909 no COG classification, a search against the KEGG orthology database (108) was performed. For
910 the major COG/KEGG categories (metabolism, cellular processes and signaling, and Information
911 storage and processing), sub-classifications are shown in C, D, and E, respectively.

912 **Figure 4**. HGT impact on AGF central metabolic abilities. Pathways for sugar metabolism are
913 highlighted in blue, pathways for amino acid metabolism are highlighted in red, pathways for
914 cofactor metabolism are highlighted in green, pathways for nucleotide metabolism are

highlighted in grey, pathways for lipid metabolism are highlighted in orange, fermentation pathways are highlighted in purple, while pathways for detoxification are highlighted in brown. The double black lines depict the hydrogenosomal outer and inner membrane. Arrows corresponding to enzymes encoded by horizontally transferred transcripts are shown with thicker dotted lines and are given numbers 1 through 46 as follows. Sugar metabolism (1-9): 1. Xylose isomerase, 2. Xylulokinase, 3. Ribokinase, 4. 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, 5. 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase, 6. Phosphoenolpyruvate synthase, 7. Aldose-1-epimerase, 8. Galactokinase, 9. Galactose-1-phosphate uridyltransferase. Amino acid metabolism (10-18): 10. Aspartate-ammonia ligase, 11. Tryptophan synthase (TrpB), 12. Tryptophanase, 13. Monofunctional prephenate dehydratase, 14. Serine-O-acetyltransferase, 15. Cysteine synthase, 16. Low-specificity threonine aldolase, 17. 5'-methylthioadenosine nucleosidase/5'-methylthioadenosine phosphorylase (MTA phosphorylase), 18. Arginase. Cofactor metabolism (19-26): 19. Pyridoxamine 5'-phosphate oxidase, 20. L-aspartate oxidase (NadB), 21. Quinolate synthase (NadA), 22. NH(3)-dependent NAD(+) synthetase (NadE), 23. 2-dehydropantoate 2-reductase, 24. dephosphoCoA kinase, 25. Dihydrofolate reductase (DHFR) family, 26. Dihydropteroate synthase. Nucleotide metabolism (27-34): 27. GMP reductase, 28. Trifunctional nucleotide phosphoesterase, 29. deoxyribose-phosphate aldolase (DeoC), 30. Oxygen-sensitive ribonucleoside-triphosphate reductase class III (NrdD), 31. nucleoside/nucleotide kinase family protein, 32. Cytidylate kinase-like family, 33. thymidylate synthase, 34. thymidine kinase. Pyruvate metabolism (fermentation pathways) (35-39): 35. D-lactate dehydrogenase, 36. bifunctional aldehyde/alcohol dehydrogenase family of Fe-alcohol dehydrogenase, 37. Butanol dehydrogenase family of Fe-alcohol dehydrogenase, 38. Zn-type alcohol dehydrogenase, 39. Fe-only hydrogenase. Detoxification reactions (40-43): 40.

938 Phosphoglycolate phosphatase, 41. Glyoxal reductase, 42. Glyoxalase I, 43. Glyoxalase II. Lipid
939 metabolism (44-46): 44. CDP-diacylglycerol--serine O-phosphatidyltransferase, 45.
940 lysophospholipid acyltransferase LPEAT, 46. methylene-fatty-acyl-phospholipid synthase.
941 Following the numbers, between parentheses, the distribution of the specific event across AGF
942 genera is shown where (all) indicates the event was detected in all 7 genera, while a minus sign
943 followed by a genus indicates that the event was detected in all but that/those genus/genera.
944 Genera are represented by letters as follows: A, *Anaeromyces*; C, *Caecomyces*; F, *Feramyces*, N,
945 *Neocallimastix*, O, *Orpinomyces*; Pe, *Pecoramyces*; Pi, *Piromyces*. Abbreviations: CDP-DAG,
946 CDP-diacylglycerol; 7,8 DHF, 7,8 dihydrofolate; EthA, ethanolamine; Gal, galactose; GAP,
947 glyceraldehyde-3-P; Glu, glucose; GSH, glutathione; I, complex I NADH dehydrogenase;
948 NaMN, Nicotinate D-ribonucleotide; Orn, ornithine; PEP, phosphoenol pyruvate; Phenyl-pyr,
949 phenylpyruvate; PRPP, phosphoribosyl-pyrophosphate; Ptd, phosphatidyl; SAM; S-
950 adenosylmethionine; THF, tetrahydrofolate.
951 **Figure 5.** (A) Maximum likelihood tree showing the phylogenetic affiliation of AGF
952 galactokinase. AGF genes highlighted in light blue clustered within the Flavobacteriales order of
953 the Bacteroidetes phylum and were clearly nested within the bacterial domain (highlighted in
954 green) attesting to their non-fungal origin. Fungal galactokinase representatives are highlighted
955 in pink. (B) Maximum likelihood tree showing the phylogenetic affiliation of AGF Fe-only
956 hydrogenase. AGF genes highlighted in light blue clustered within the Thermotogae phylum and
957 were clearly nested within the bacterial domain (highlighted in green) attesting to their non-
958 fungal origin. *Stygiella incarcerata* (anaerobic Jakobidae) clustered with the Thermotogae as
959 well, as has recently been suggested (89). Fe-only hydrogenases from *Gonopodya prolifera*
960 (Chytridiomycota) (shown in orange text) clustered with the AGF genes. This is an example of

961 one of the rare occasions (n=24) where a non-AGF basal fungal representative showed an HGT
962 pattern with the same donor affiliation as the Neocallimastigomycota. Other basal fungal Fe-only
963 hydrogenase representatives are highlighted in pink and clustered outside the bacterial domain.
964 (C) Maximum likelihood tree showing the phylogenetic affiliation of AGF L-aspartate oxidase
965 (NadB). AGF genes highlighted in light blue clustered within the Delta-Proteobacteria class and
966 were clearly nested within the bacterial domain (highlighted in green) attesting to their non-
967 fungal origin. As de-novo NAD synthesis in fungi usually follow the five-enzyme pathway
968 starting from tryptophan, as opposed to the two-enzyme pathway from aspartate, no NadB were
969 found in non-AGF fungi and hence no fungal cluster is shown in the tree. (D) Maximum
970 likelihood tree showing the phylogenetic affiliation of AGF oxygen-sensitive ribonucleotide
971 reductase (NrdD). AGF genes highlighted in light blue clustered with representatives from
972 Candidate phylum Dependentiae and were clearly nested within the bacterial domain
973 (highlighted in green) attesting to their non-fungal origin. Fungal NrdD representatives are
974 highlighted in pink. GenBank accession numbers are shown in parentheses. Alignment was done
975 using the standalone MAFFT aligner (69) and trees were constructed using IQ-tree (70).
976 **Figure 6**. HGT in the AGF CAZyome shown across the seven genera studied. Glycosyl
977 Hydrolase (GH), Carboxyl Esterase (CE), and Polysaccharide Lyase (PL) families are shown to
978 the left. The color of the cells depicts the prevalence of HGT within each family. Red indicates
979 that 100% of the CAZyme transcripts were horizontally transferred. Shades of red-orange
980 indicate that HGT contributed to > 50% of the transcripts belonging to that CAZy family. Dark
981 blue indicates that 100% of the CAZyme transcripts were of fungal origin. Shades of blue
982 indicate that HGT contributed to < 50% of the transcripts belonging to that CAZy family. The
983 numbers in each cell indicate the affiliation of the HGT donor as shown in the key to the right.

984 **Figure 7**. Principal-component analysis biplot of the distribution of CAZy families in AGF
985 genomes (★), compared to representatives of other basal fungi belonging to the
986 Mucoromycotina (⬣), Chytridiomycota (○), Blastocladiomycota (■), Entomophthoromycotina
987 (⬭), Mortierellomycotina (⬡), Glomeromycota (✚), Kickxellomycotina (□), and
988 Zoopagomycotina (✖). CAZy families are shown as colored dots. The color code used was as
989 follows: green, CAZy families that are absent from AGF genomes; black, CAZy families present
990 in AGF genomes and with an entirely fungal origin; blue, CAZy families present in AGF
991 genomes and for which HGT contributed to < 50% of the transcripts in the examined
992 transcriptomes; red, CAZy families present in AGF genomes and for which HGT contributed to >
993 50% of the transcripts in the examined transcriptomes. The majority of CAZyme families
994 defining the AGF CAZyome were predominantly of non-fungal origin (red and blue dots).

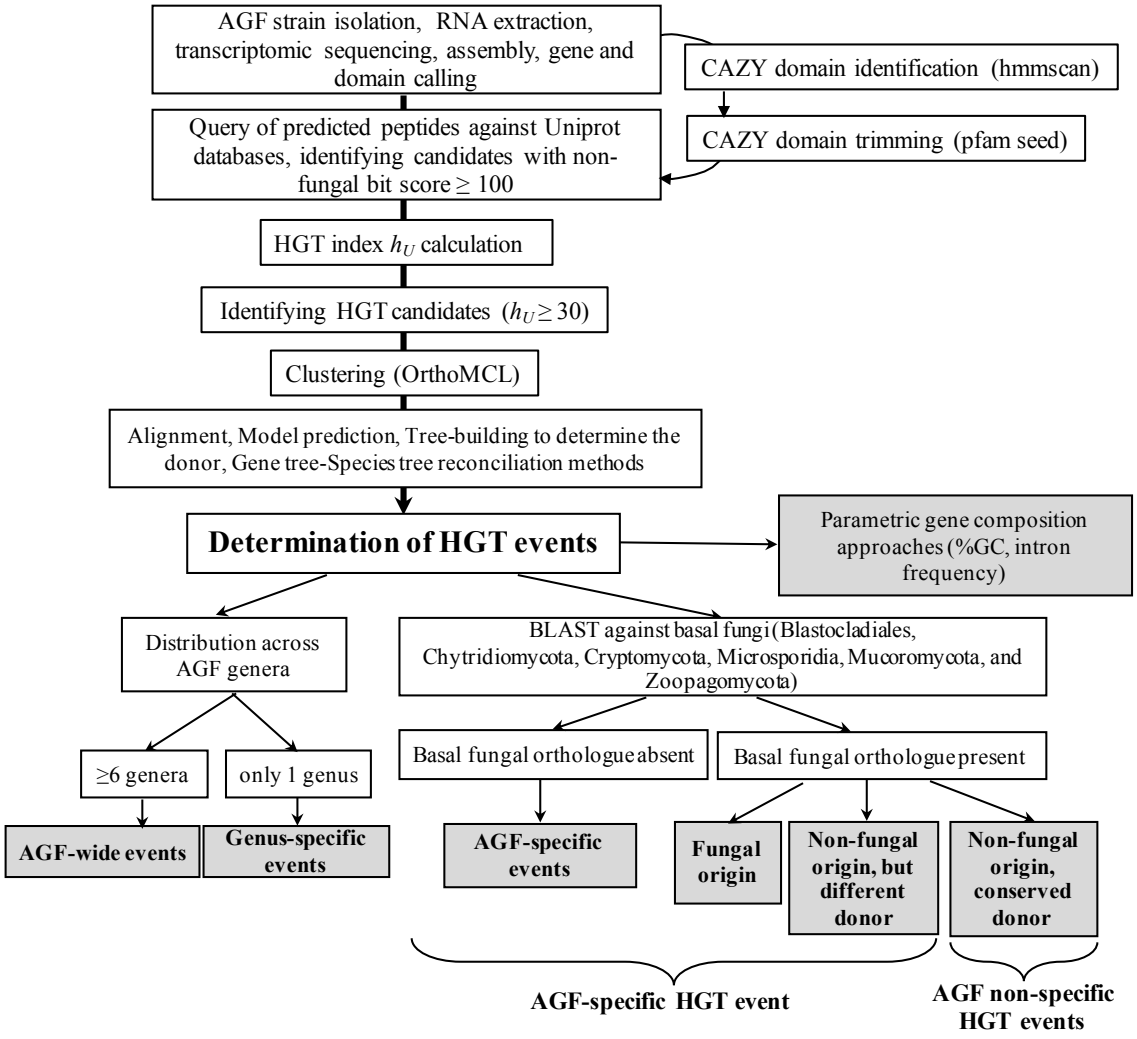
Table 1: Neocallimastigomycota strains analyzed in this study.

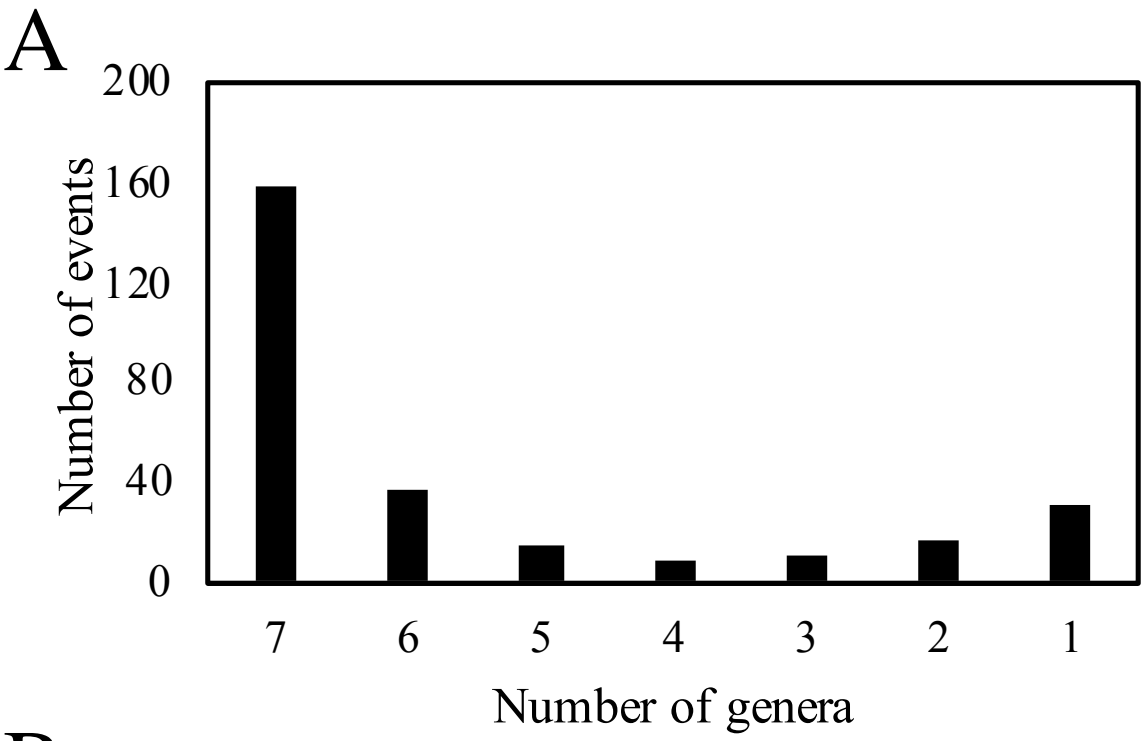
Genus	Species	Strain	Host	Isolation source	Location	LSU Genbank accession number	Reference
Anaeromyces	*contortus*	C3G	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121936	This study
Anaeromyces	*contortus*	C3J	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121942	This study
Anaeromyces	*contortus*	G3G	Goat (*Capra aegagrus hircus*)	Feces	Stillwater, OK	MF121935	This study
Anaeromyces	*contortus*	Na	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121943	This study
Anaeromyces	*contortus*	O2	Cow (*Bos taurus*)	Feces	Stillwater, OK	MF121931	This study
Anaeromyces	*robustus*	S4	Sheep (*Ovis aries*)	Feces	Santa Barbara, CA	NA*	(45)
Caecomyces	sp.	Iso3	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992499	This study
Caecomyces	sp.	Brit4	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992500	This study
Feramyces	*austinii*	F2c	Aoudad sheep (*Ammotragus lervia*)	Feces	Stillwater, OK	MG605675	This study
Feramyces	*austinii*	F3a	Aoudad sheep (*Ammotragus lervia*)	Feces	Stillwater, OK	MG584226	This study
Neocallimastix	*californiae*	G1	Goat (*Capra aegagrus hircus*)	Feces	Santa Barbara, CA	Genomic sequence**	(45)
Neocallimastix	cf. *cameroonii*	G3	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992493	This study
Neocallimastix	cf. *frontalis*	Hef5	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992494	This study
Orpinomyces	cf. *joyonii*	D3A	Cow (*Bos taurus*)	Digesta	Stillwater, OK	MG992487	This study

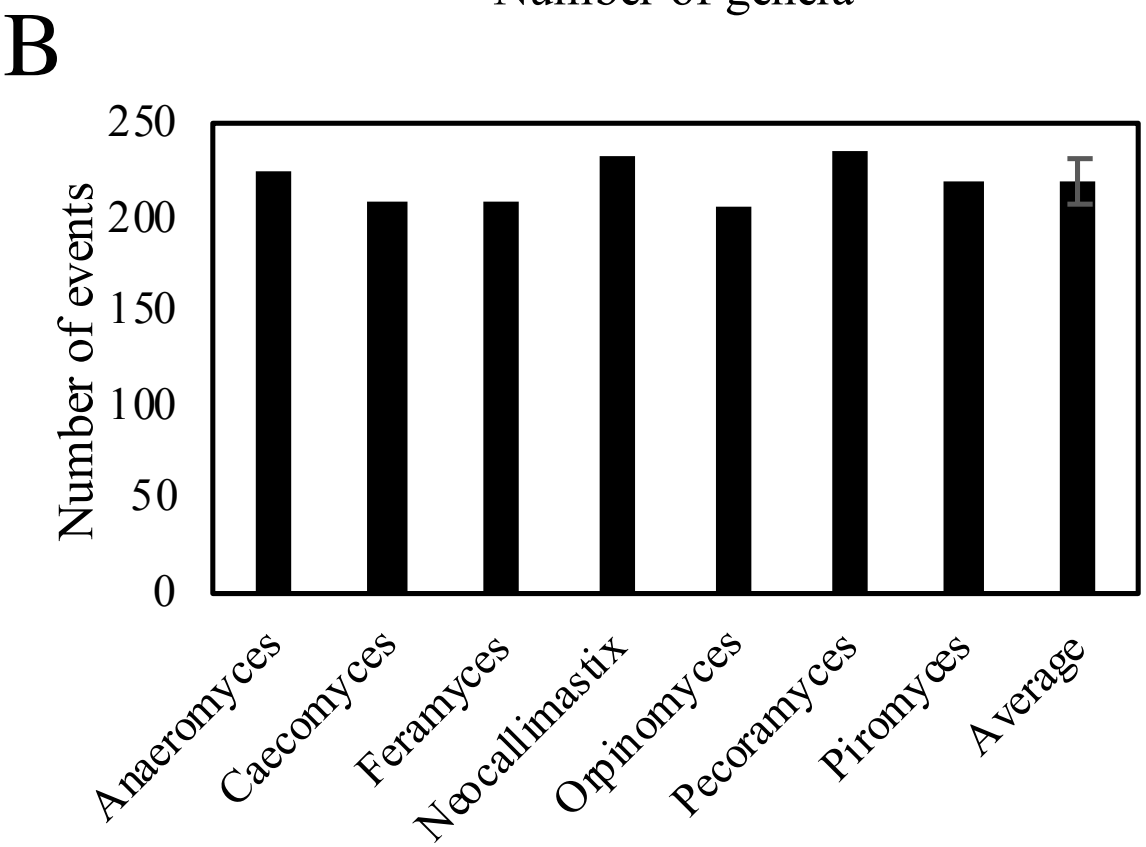
Orpinomyces	cf. *joyonii*	D3B	Cow (*Bos taurus*)	Digesta	Stillwater, OK	MG992488	This study
Orpinomyces	cf. *joyonii*	D4C	Cow (*Bos taurus*)	Digesta	Stillwater, OK	MG992489	This study
Pecoramyces	*ruminantium*	C1A	Cow (*Bos taurus*)	Feces	Stillwater, OK	JN939127	(42, 55)
Pecoramyces	*ruminantium*	S4B	Sheep (*Ovis aries*)	Feces	Stillwater, OK	KX961618	This study
Pecoramyces	*ruminantium*	FS3C	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992492	This study
Pecoramyces	*ruminantium*	FX4B	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992491	This study
Pecoramyces	*ruminantium*	YC3	Cow (*Bos taurus*)	Rumen	Stillwater, OK	MG992490	This study
Piromyces	*finnis*	finn	Horse (*Equus caballus*)	Feces	Santa Barbara, CA	Genomic sequence**	(45)
Piromyces	sp.	A1	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992496	This study
Piromyces	sp.	A2	Sheep (*Ovis aries*)	Feces	Stillwater, OK	MG992495	This study
Piromyces	sp.	B4	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992497	This study
Piromyces	sp.	B5	Cow (*Bos taurus*)	Feces	Stillwater, OK	MG992498	This study
Piromyces	sp.	E2	Indian Elephant (*Elephas maximus*)	Feces	London, UK	NA	(45, 106)

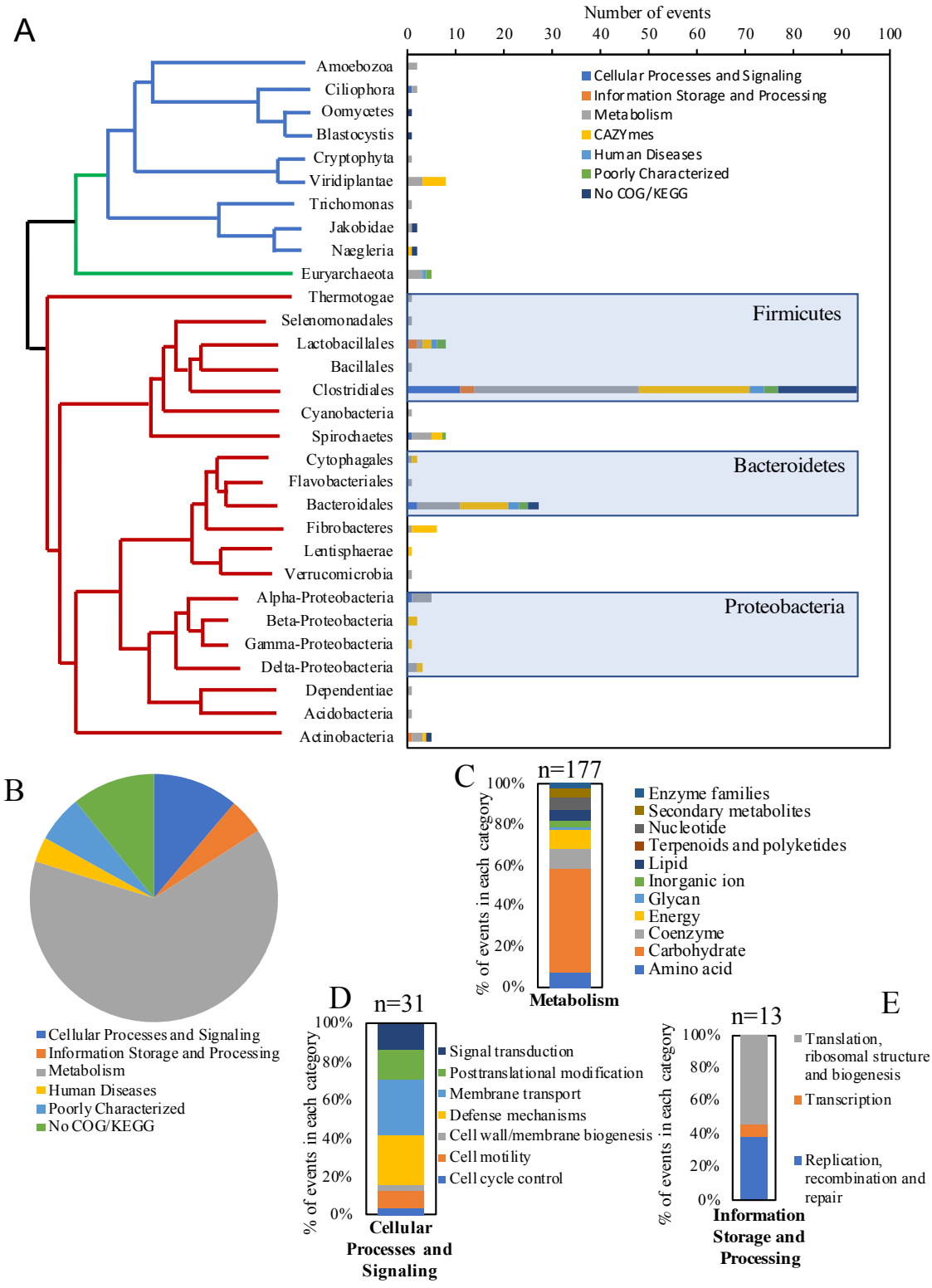
*NA: Not available

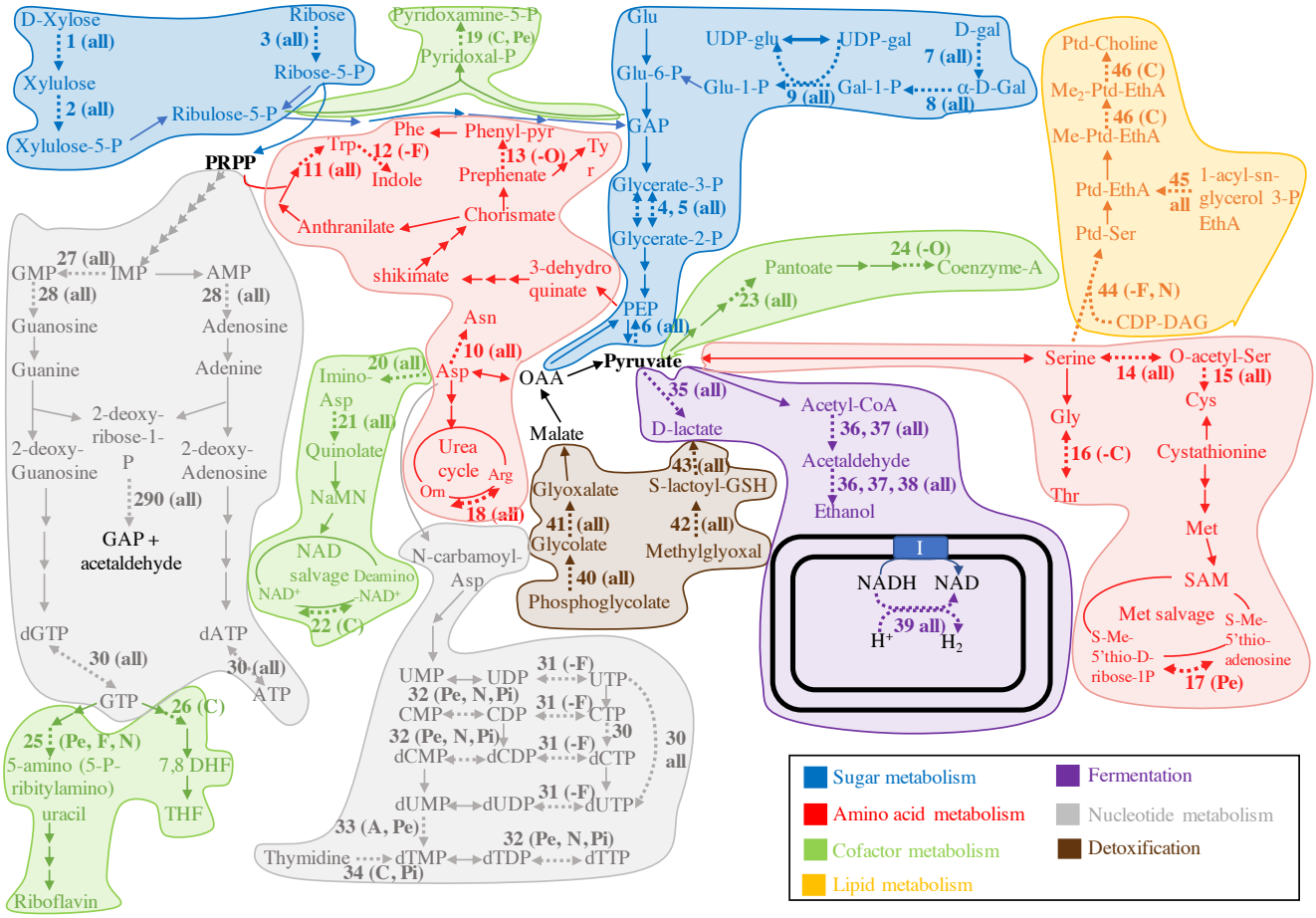
** LSU sequence was extracted from the genomic assembly. No LSU accession number was available.



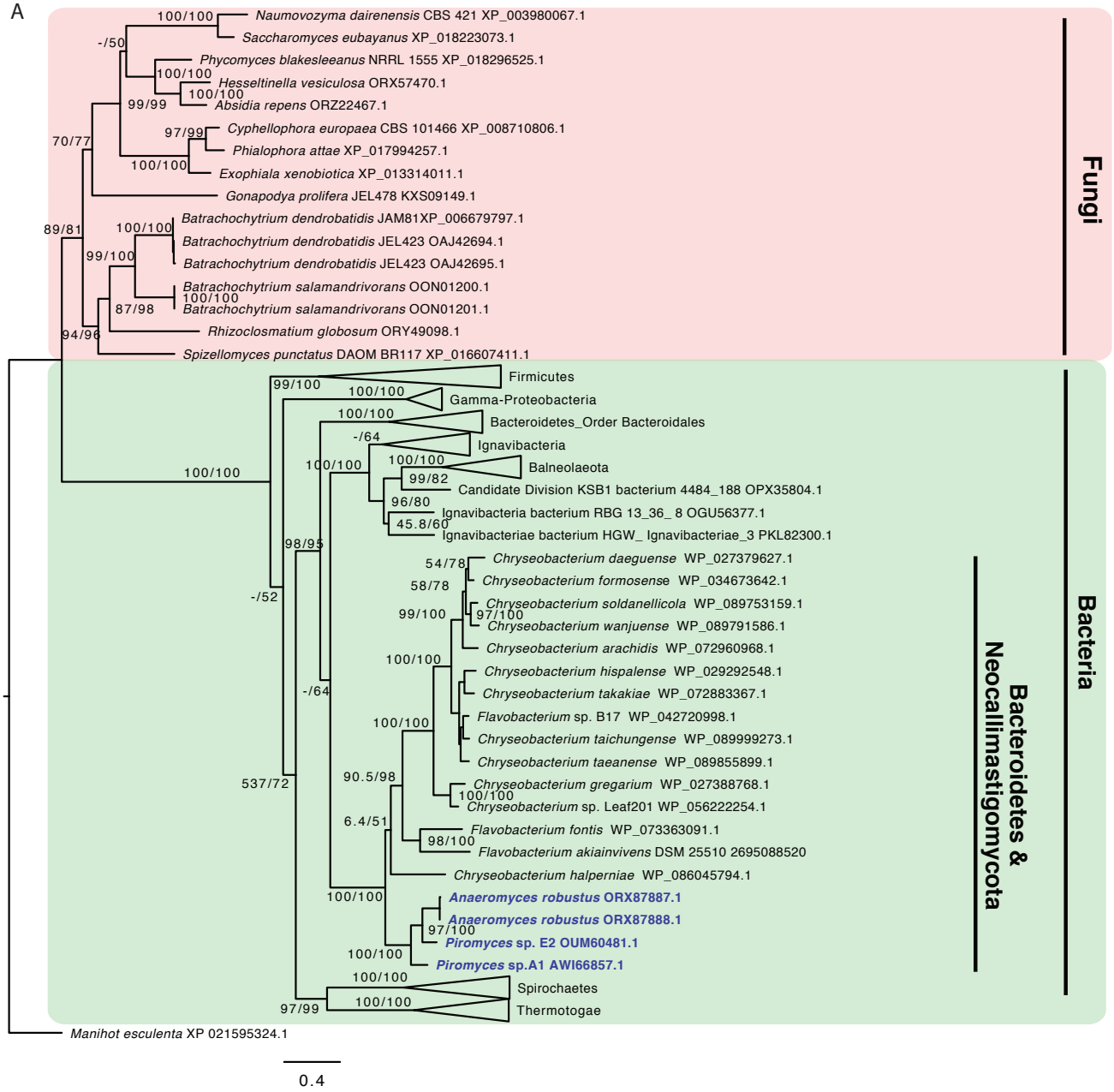




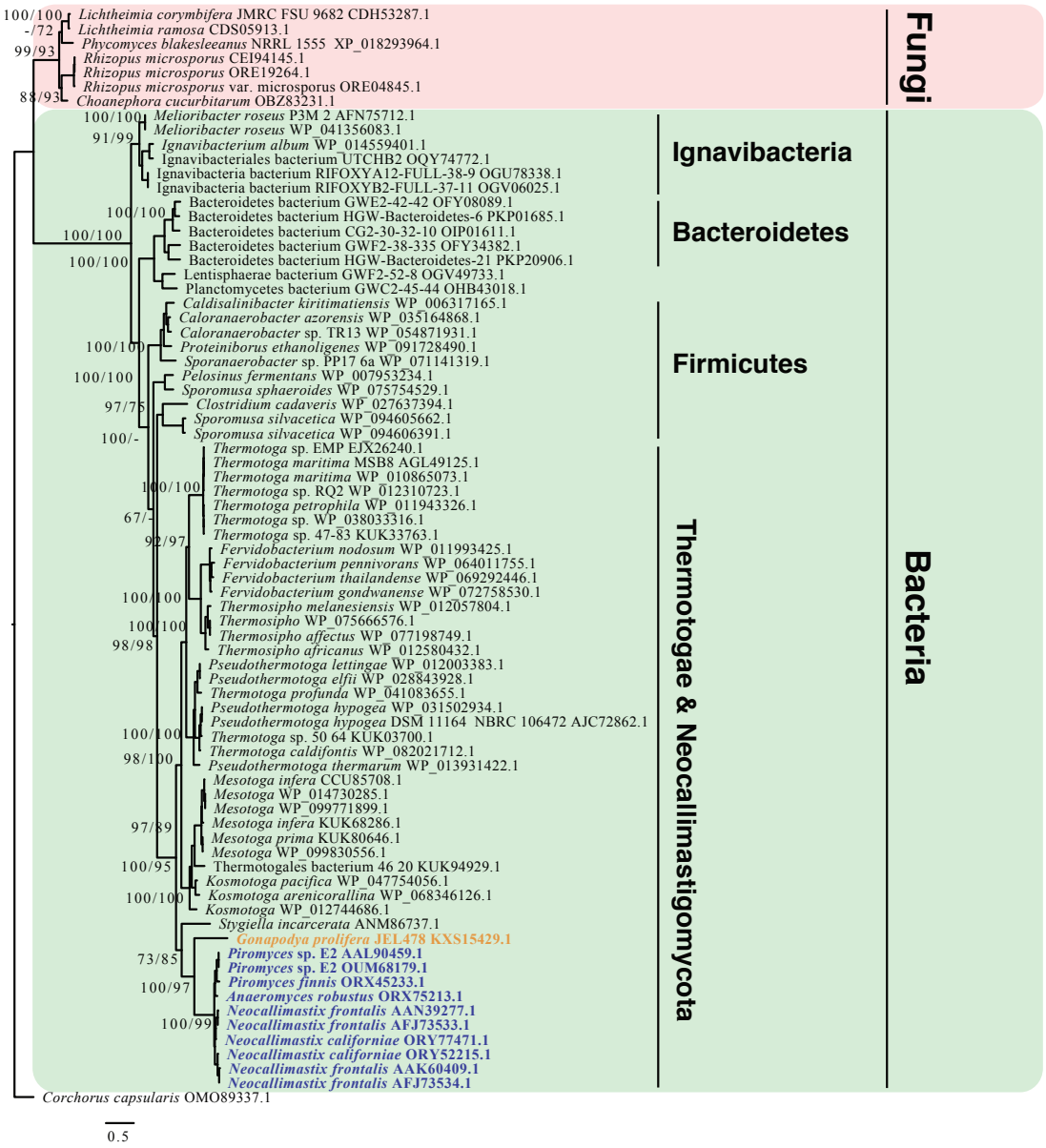




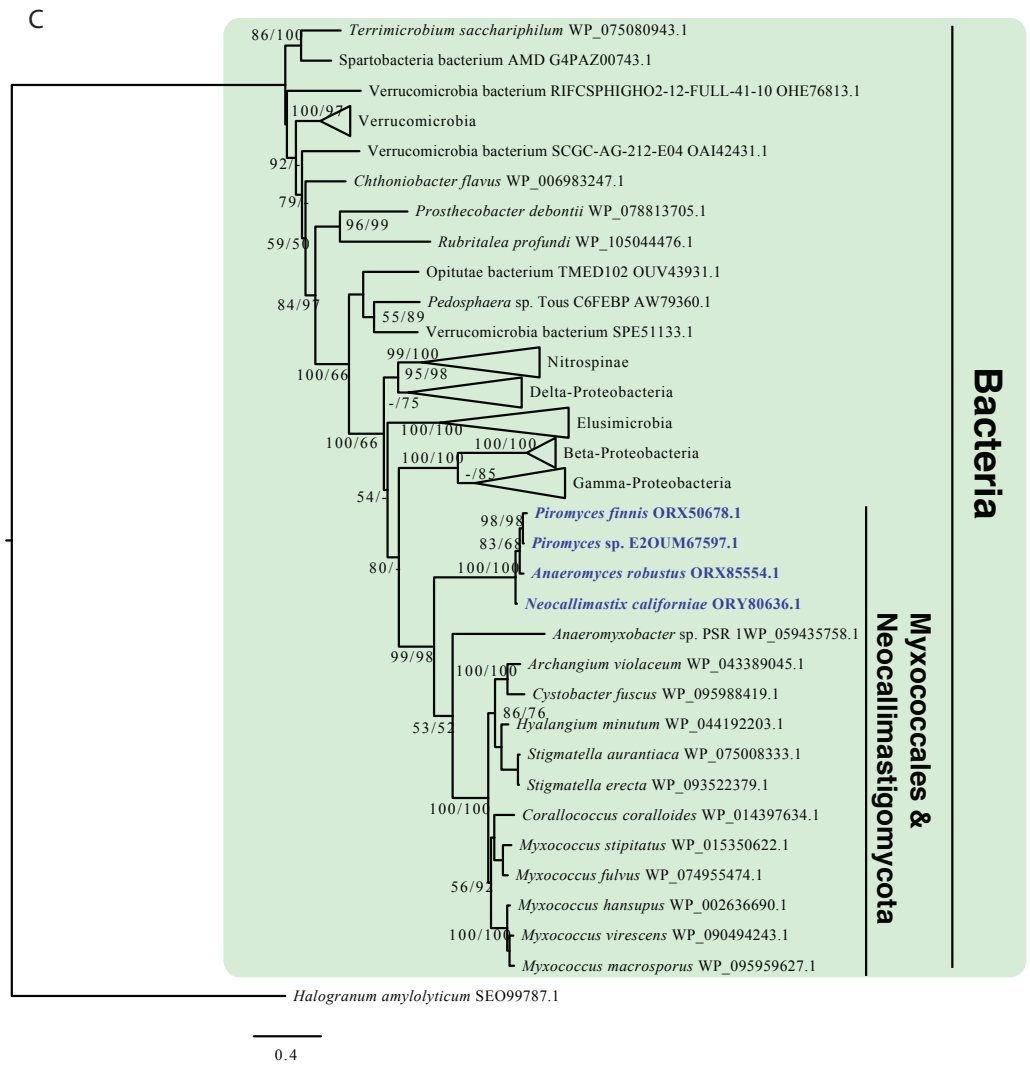
A

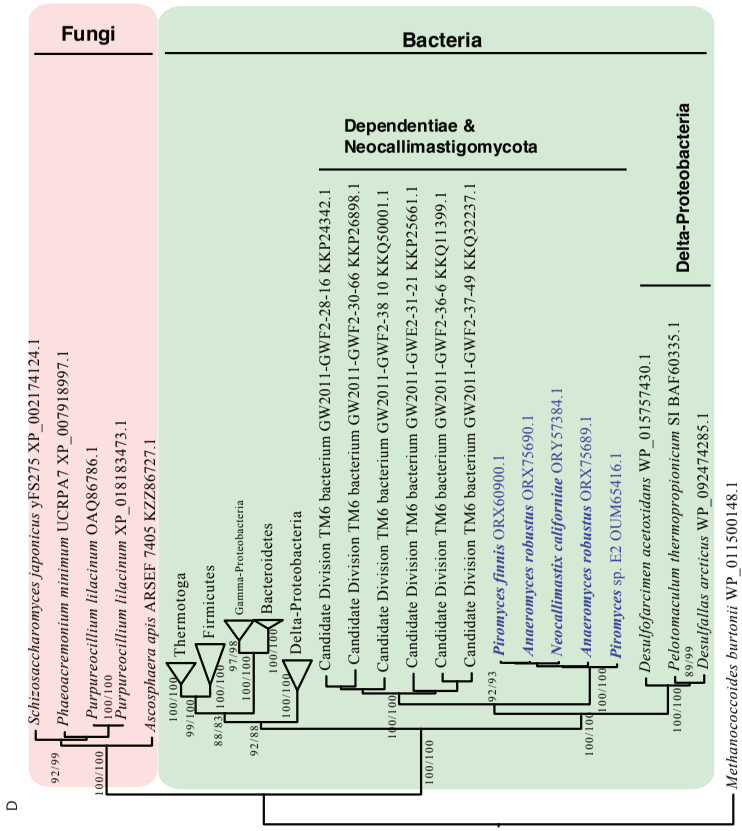


B



Applied and Environmental Microbiology





Family	Genus						
	Anaeromyces	*Caecomyces*	*Pecoramyces*	*Piromyces*	*Neocallimastix*	*Feramyces*	*Orpinomyces*
GH1	8						4
GH2	4		4	4	4	4	4
GH3	7, 15	7, 15	7, 15	7, 15	7		7
GH5	12	4, 12	4, 12	4, 12	4, 12	4, 12	4, 12
GH6	10	10	10	10	10	10	10
GH8	3, 5	3	3, 5	3, 5	3, 5	5	3, 5
GH9							
GH10	4, 12	4, 12	4, 12	4, 12	4, 12	4, 12	4, 12
GH11	3, 4	3, 4	3, 4	3, 4	3, 4	3, 4	3, 4
GH13	2, 4, 6	2, 4, 6	2, 4, 6	2, 6	2, 4, 6	2, 4, 6	2, 4, 6
GH16	2, 4		2	4	2, 4, 13	2, 4	2, 4
GH18		12	12	12	12	12	
GH20							
GH24	4, 9			4			9
GH25	4	4	4	4	4	4	
GH26	3	3	3	3	3		
GH28			2		2, 13		2
GH30	4				4	4	
GH31			1				
GH32	5		5		2, 13	5	5
GH36							
GH37							
GH39	2, 4	4	2, 4	4	2, 4		4
GH43	4	4, 12	4	4, 8, 12	4, 12	4, 11	4
GH45							
GH47	15		15	15	15	15	
GH48	12	12	12	12	12	12	12
GH53	4	4	4	4	4	4	4
GH57							
GH64	4	4	4		4	4	
GH67	2	2	2	2			
GH76						2	
GH78			12				
GH88			12	12	12	12	12
GH95	2						2
GH97					3	3	
GH108				14			
GH114							
GH115	4	4	4	4	4	4	
CE1	3, 4	3, 4	3, 4	3, 4	3, 4	3, 4	3, 4
CE2	6, 12	6, 12	6, 12	6, 12	6, 12	2, 6	6
CE3	4	4	4	4	4	4	4
CE4	3	3	3	3	3	3	3
CE6	4	4	4	4	4	4	4
CE7	4						
CE8	13	13	13	13	13	13	13
CE12	4, 12		4	4, 12	4	2, 4	4
CE15	12	3, 12	12	12	3, 12	12	12
CE16					13		
PL1							
PL3							
PL4	12	12	12	12	12	12	12
PL9	4, 11	4	4, 11	7, 11	4, 11		11, 14
PL11					4		

Donor Key			
1	Actinobacteria		Bacteria
2	Bacteroidetes		
3	Fibrobacter		
4	Clostridiales	Firmicutes	
5	Lactobacillales		
6	Unclassified Firmicutes		
7	Lentisphaerae		
8	Beta-Proteobacteria	Proteobacteria	
9	Gamma-Proteobacteria		
10	Delta-Proteobacteria		
11	Spirochaetes		
12	Bacteria (unnested)		
13	Viridplantae		Eukaryota
14	Neagelaria		
15	Unclassified Eukaryotes